DESIGN, SYNTHESIS, AND ANALYSIS OF CONJUGATED ACTIVATORS FOR MODULATION OF ALTERNATIVE SPLICING

BY

ADAM DAVID LANGENFELD

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Associate Professor Anne Baranger, Chair
Professor John Katzenellenbogen
Professor Scott Silverman
Professor Wilfred van der Donk

ABSTRACT

Recent genome wide analysis has estimated that >95% of human genes are alternatively spliced, allowing for the production of many protein isoforms from a single gene. Splicing is controlled by SR proteins, factors which bind to exonic splicing enhancers (ESEs) through RNA binding domains and recruit the splicosome via protein-protein interactions mediated by their RS domains, protein domains rich in highly charged arginine-serine dipeptides. Defects in alternative splicing have been linked to several diseases, including Cystic Fibrosis, breast cancer, and Spinal Muscular Atrophy (SMA). SMA is a neurodegenerative disorder caused by the loss of survival of motor neuron 1 (*SMN1*) gene. A nearly identical copy of the gene, *SMN2*, encodes an identical protein but contains a C→T transition on the ESE of exon 7, disrupting SR protein binding and resulting in substantial exon 7 skipping.

Significant research efforts have focused on redirecting the splicing pattern of *SMN2* to produce a full-length SMN protein normally generated by *SMN1*. Splicing modulators have ranged from small molecule activators such as sodium phenylbutyrate and valproic acid, to viral vectors encoding antisense oligonucleotides. In particular, a synthetic peptide-oligonucleotide conjugate containing a short chain of arginine-serine dipeptides to mimic the RS domain has been shown to redirect splicing patterns. However, little work has been done to explore the potential flexibility of the design of this synthetic RS domain. This dissertation reports efforts toward the design and synthesis of several synthetic SR protein mimics and the experimental design to determine their ability to increase inclusion of exon 7 in *SMN2*. The synthetic RS domains consist of three motifs – α-peptide, β-peptide, and peptide dendrimer – attached to antisense oligonucleotides targeted to *SMN2* exon 7. The synthesis and conjugation of these mimics is straightforward, allowing for easy derivatization. The efficacy of these molecules was examined using an *in vitro* splicing assay. Analysis of these molecules using *in vitro* splicing assays indicates that neither synthetic RS domains nor antisense oligonucleotides alone affect splicing patterns, but conjugated molecules may play a beneficial role.

ACKNOWLEDGEMENTS

I would like to thank a number of people who have provided me support and guidance throughout this project. I would first like to thank my research advisor, Dr. Anne Baranger, for giving me the freedom to explore the possibilities of this project while offering guidance and suggestions to help keep me moving in the right direction. I would also like to thank my doctoral thesis committee members, Professor John Katzenellenbogen, Professor Scott Silverman, and Professor Wilfred van der Donk, for providing useful advice and suggestions at various points throughout my career. I would especially like to thank Dr. K for acting as my on-campus research advisor for the past year and providing additional insights as I finished my research.

I would like to thank all past and present members of the Baranger lab for their thoughtful advice and instruction on different laboratory techniques, as well as helpful discussions on improving my project. I would especially like to thank Stacie Richardson for her great advice and suggestions on all aspects of the project, her helpful tips for laboratory techniques, and for making sure that I was pushing ahead even when I was not confident in myself.

I am grateful for the love and support I received from my family and friends during the course of this work. I am especially grateful for the advice and wisdom of my mother Michelle, who pushed me to work hard and advocate for myself, and for my father David, who taught me to be patient in the face of adversity. Finally, I am most grateful for my loving wife Emily, who has been the most supportive of all and did not allow me to give up even when I was tempted to do so. Thank you for helping me get here.

TABLE OF CONTENTS

CHAPTER 1:

ALTERNATIVE SPLICING AND DISEASE

1.1 Alternative Splicing

Alternative splicing is a process by which non-coding regions are removed from messenger RNA (mRNA) to produce a variety of protein products. The frequency of alternative splicing is high in the human genome, with recent estimates indicating that greater than 95% of human genes are spliced in some fashion.[1-3] The study of alternative splicing and its mechanisms has produced a number of insights into the required components that are used to process mRNA. These components are separated into two categories based on their proximity to the mRNA. *Cis*-acting elements are sequence motifs located within the mRNA that can direct splicing proteins to the correct gene encoding regions.[4,5] *Trans*-acting elements, on the other hand, are the protein components that bind to the mRNA and control the splicing mechanism by assisting in the recognition of exons.[6,7] When combined, these elements control the choice of gene-encoding sequences that will be included in the mature mRNA, and direct the reactions to produce the final product.

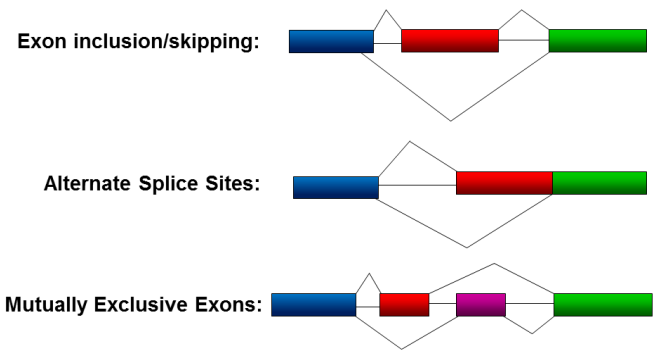


Figure 1.1. Possible outcomes of alternative splicing

The mechanism of alternative splicing involves a series of steps that remove the non-coding introns from the exons, producing a mature mRNA that can be used as a template for translation. It involves recognition of the RNA sequence elements involved in splicing by a number of proteins that constitute the spliceosome, assembly of different protein elements onto the pre-mRNA, and consecutive trans-esterification reactions that remove the introns and connect the exons together. This series of reactions is highly complex, requiring a number of splicing proteins and RNA molecules that must attach and detach from the core splicing proteins on the mRNA. The outcome of the splicing reaction can lead to multiple mRNA products through differential exon choices. Possible outcomes include removal or

inclusion of a particular exon, choice of alternative splice sites that may or may not lead to inclusion of a particular exon, or use of mutually exclusive exons that can lead to two very different mRNA products (**Figure 1.1**).

Unlike constitutive splicing, alternative splicing relies on additional sequence elements in order to strengthen recognition of the pre-mRNA by the spliceosome and ensure that the correct exon is inserted into the mature mRNA. The relative strength of these elements can be regulated, so that certain exons can be included or excluded depending on the conditions of the splicing event. Weaker splicing sequences are flanked by additional sequence elements to ensure that they are recognized by the spliceosome. These elements are called splicing enhancer sequences, and they can be located either within the exon (exonic splicing enhancer – ESE) or in the intron (intronic splicing enhancer – ISE). In contrast, additional sequences called splicing silencers (ESS or ISS) are present within the pre-mRNA, acting as negative regulators of splicing and preventing exon inclusion. These sequence elements are particularly useful if an exon is not included in the mature mRNA under normal circumstances.

In addition to *cis*-acting elements, *trans*-acting elements play an important role in alternative splicing. A family of proteins called the SR proteins, named for their intrinsically disordered RS domains rich in arginine-serine dipeptides, act as positive regulators of alternative splicing by binding to enhancer sequences on the pre-mRNA and recruiting the spliceosome to that sequence (**Figure 1.2a)**.[7,8] These proteins are modular, containing the RS domain and an RNA binding domain that targets the enhancer sequences. Research into the structure and function of the SR proteins indicates that the RS domain is very general, and can be exchanged with different RS domains of other SR proteins and retain the activity of the original protein.[9] In contrast, a class of proteins called the heterogenous nuclear ribonucleoproteins (hnRNPs) act as negative regulators of alternative splicing by targeting the silencing elements on the pre-mRNA and preventing interaction of the particular exon with the splicing machinery (**Figure 1.2a**).[10] Although some interactions of hnRNPs with pre-mRNA may be promote splicing, including the promotion of exon 7 inclusion in *SMN2* mRNA through interaction of hnRNP-G with SR-like protein Htra2β1,[11] in general the main function of the hnRNPs is to block binding of pre-mRNA by the spliceosome.

1.2 SR Proteins

SR proteins play an important role in splicing, and their properties can be utilized to develop splicing modifiers. The structure of SR proteins is modular, consisting of one or two N-terminal RNA binding domains (RBD) and a C-terminal intrinsically disordered RS (arginine-serine rich) domain that functions primarily as a protein-protein interaction domain, but has also been proposed to interact with RNA.[12-15] In alternative splicing, an SR protein binds to an ESE via its RBD (**Figure 1.2b**). It then interacts with the small subunit of SR-like protein U2AF ($U2AF^{35}$), resulting in recruitment of the large subunit, $U2AF^{65}$, to the polypyrimidine tract of the adjacent intron. This complex then recruits the remainder of the spliceosome to the mRNA.[6,16]

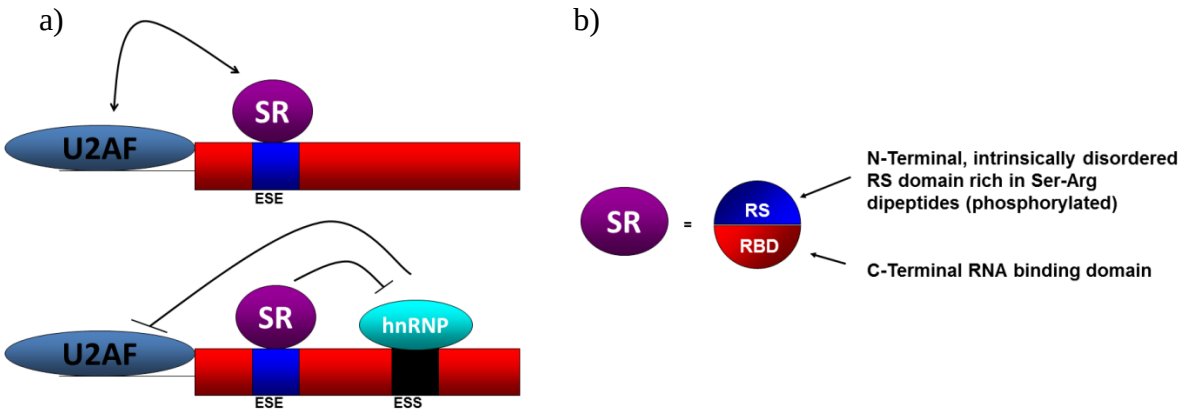


Figure 1.2. Structure and roles of SR proteins. a) SR proteins bind to ESE motifs on pre-mRNA and recruit spliceosomal proteins while inhibiting binding of hnRNP molecules. b) SR proteins consist of two domains.

The determinants of RS domain strength, defined as the ability of the domain to activate splicing, were systematically studied by Graveley and co-workers in 1998.[17] The strength of 5 different RS domains was found to be directly correlated with RS content, specifically the number of RS tetrapeptides, although other factors may be involved. Despite this correlation, the sequence requirements to activate alternative splicing are not clear, although they are less stringent than for constitutive splicing. For example, non-charged amino acids, such as glycine, can be substituted for charged residues in SR proteins without affecting activation.[18,19] In small peptide models, however, the charged residues are essential for activity.[20] Some of the

variability of the effects of alternative splicing modulators may arise from differences in the strength of the splicing signals present in the pre-mRNA.

The serine residues in the RS domain are reversibly phosphorylated.[6,21] The negative charges are proposed to serve three functions: assistance in protein-protein interactions, prevention of non-specific RNA binding, and nuclear transport.[6,22,23] Manley and co-workers demonstrated in 1998 that dephosphorylation of serine is not required for the SR protein SF2/ASF in the activation of alternative splicing.[23,24] Phosphorylation was shown to enhance binding of SF2/ASF to both the U1snRNP complex and U2AF[35].[24] Increasing the salt concentration in binding studies led to a decrease in binding, demonstrating that the binding is largely ionic in nature. Phosphorylation also aids in preventing non-specific RNA binding through charge repulsion with the RNA backbone.[23] In addition, phosphorylation of serine is essential for the transport of SR proteins to the nucleus. Tarn and co-workers demonstrated that a nuclear transport protein, transportin-SR2, specifically interacts with phosphorylated RS domains and facilitates their nuclear import.[25]

In 2003, Krainer and co-workers demonstrated that a minimal RS domain composed of 10 RS (phosphorylated), RD, or RE dipeptide repeats could enhance exon 7 inclusion in *SMN2* mRNA. Synthetic RS domains without phosphorylated serine (or a negatively charged equivalent) could not activate *SMN2* splicing.[20] Caputi and co-workers demonstrated similarly that a minimal RS domain comprised of 8 dipeptide repeats was functional as a modulator of splicing in the *Bcl-x* gene both *in vitro* and *in vivo*.[26] Together, this work provides two important insights about the requirements for synthetic SR proteins. First, a synthetic RS domain can be permanently and fully "phosphorylated". Second, synthetic RS domains can be smaller than the natural counterpart (SF2/ASF, for example, contains 15 RS dipeptide repeats) and still activate splicing *in vitro* and, importantly, *in vivo*.

1.3 Alternative Splicing and Disease

The high prevalence of alternative splicing in the human genome increases the potential for errors in the splicing mechanism that can lead to a disease phenotype. Malfunction of the splicing mechanism can lead to gain or loss of particular sequence elements that can affect the processing of mRNA or production of dysfunctional protein, which can have profoundly

negative effects on human health.[27-29] Recent estimates indicate that as high as 50% of human genetic disease can be directly linked so aberrant splicing patterns by defining splicing-related as any disease state in which some part of the splicing mechanism is malfunctioning.[30,31] There are many possible ways in which alternative splicing can malfunction. Mutations in *cis*-acting elements on the pre-mRNA can lead to loss of enhancer sequences, gain of silencing sequences, activation of cryptic (rarely used) splicing sequences, or changes in secondary structure, leading to undesired splicing patterns.[32,33] In addition, mutations to genes encoding *trans*-acting elements can have more global effects on alternative splicing.[7]

Table 1.1 Examples of Splicing-Related Diseases[28]

Disease	**Gene/Splicing Factors Affected**	**Genetic Effect of Mutation**
Duchenne Muscular Dystrophy	*Dystrophin*	Mutation of dystrophin protein
Myotonic Dystrophy	*DMPK*, MBNL-1	CUG-repeat expansion; sequestration of MBNL-1
Cystic Fibrosis	*CFTR*	Loss of CFTR protein
Spinal Muscular Atrophy	*SMN1*	Loss of SMN protein

Many examples exist of diseases that are directly related to changes in alternative splicing (**Table 1.1**).[28] Each disease has its own set of mechanisms that researchers are attempting to address, with the hope of developing effective molecular therapies. Diseases such Duchenne Muscular Dystrophy contain mutations that affect the inclusion of exons, leading to loss of functional protein. Researchers have developed treatments utilizing molecules that lead to skipping of exon 23 of *dystrophin*, redirecting the splicing of mRNA and producing more truncated, yet functional, protein, leading to the milder form of the disease.[34] Antisense molecules that cause exon skipping are currently being utilized in clinical trials for DMD.[35] Other diseases such as Myotonic Dystrophy are caused by sequestration of the muscleblind-like-1 (MBNL1) splicing factor by a repeat-expansion in the 3'-untranslated region of a different RNA, and researchers developing molecules to free MBNL by binding the RNA secondary structure, allowing it to function as a splicing factor in other systems.[36,37] Spinal Muscular

Atrophy (SMA) is a well-characterized splicing-related disease that is also amenable to therapies that aim to change alternative splicing and protein production.

1.4 Spinal Muscular Atrophy

Spinal muscular atrophy (SMA) is a pediatric, autosomal recessive neurodegenerative disorder that is characterized by degeneration of alpha motor neurons in the anterior horn of the spinal cord. This motor degeneration leads to proximal muscle weakness, pulmonary issues, and inability to sit, stand, or walk. SMA is separated into three clinical subtypes based on disease severity.[38] Type I (Werdnig–Hoffmann disease) is the most severe subtype. Patients are diagnosed within the few months of life and show many of the characteristic symptoms of the disease, usually dying before the age of 2. Type II (Dubowitz disease), the intermediate subtype, is usually diagnosed after the age of 2 and presents fewer, less severe variants of the symptoms. Patients are generally able to sit unassisted, but are unable to stand or walk and must be confined to a wheelchair, and generally survive into adulthood. Type III (Kugelberg–Welander disease), the least severe subtype, has a much later onset and is characterized by mild symptoms and a relatively normal lifespan. The disease rate is 1 in 6000 live births, with a carrier rate of 1 in 40. It is currently the most common genetic cause of infant mortality in the United States.

SMA is caused by the deletion or mutation of the telomeric survival-of-motor-neuron-1 (*SMN1*) gene on chromosome 5q13.[39] A nearly identical gene, *SMN2*, is present as an inverted repeat in the centromere. The two genes are 99% homologous, containing 11 translationally silent nucleotide transitions. Both genes encode for the survival-of-motor-neuron (SMN) protein, a 294-amino acid containing protein that serves several functions including snRNP biogenesis and assembly of these proteins into aggregates called gems.[40,41] The key nucleotide difference is a C→T nucleotide transition at position +6 of exon 7 of *SMN2*. This difference causes substantial skipping of exon 7 during splicing (**Figure 1.3**).[42,43] As a result, the protein that is predominantly produced by *SMN2* is truncated and quickly degraded. It has been estimated that the *SMN2* gene produces roughly 10-20% of the protein that is produced normally by a functional *SMN1* gene. Although the SMN protein is ubiquitously expressed and thus deficient in all patient tissues, only the motor neurons of the spinal cord are degraded as a result of this disease. It has been hypothesized that the small amount of functional protein produced by *SMN2* is adequate for all cell types except motor neurons, or that the SMN protein may play an

additional, unknown role in motor neurons. In particular, studies indicate that SMN protein may play an important role in axon biogenesis, which is reduced substantially in the absence of *SMN1*.[44]

The effect of the C→T transition has been debated in the literature. In 2002, it was proposed that the transition disrupts the consensus sequence of an SF2/ASF dependent ESE.[45] SF2/ASF is an SR protein that assists in splicing regulation. At the same time, a second model was proposed in which the transition instead creates an ESS at the same position, leading to increased binding of hnRNP A1, a splicing inhibitor.[46] Later work by Krainer and co-workers further explored these changes, finding that the loss of the ESE is followed by an increased ability for hnRNP A1 to non-specifically bind to *SMN2* mRNA.[47] This work indicates that a balance exists between positive and negative regulatory factors, and that the C→T transition shifts the balance toward exon exclusion.

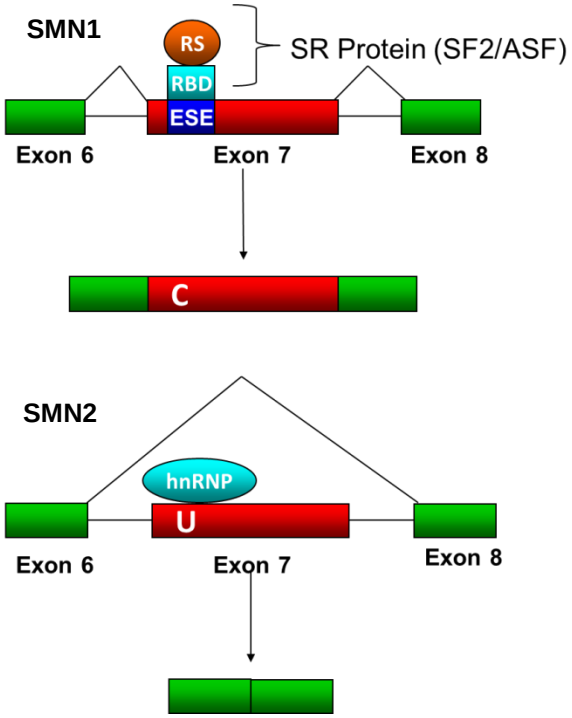


Figure 1.3. *SMN1* contains an ESE that binds SF2/ASF, leading to exon 7 inclusion. *SMN2* has a C→U transition, which disrupts this ESE, causing exon 7 exclusion.

Differences in disease severity between different SMA subtypes have led researchers to examine what causes these differences on the genetic level. Examination of patient tissues has indicated that there is variation in the number of *SMN2* gene copies that inversely correlates with disease severity – an increased number of *SMN2* copies correlates to a decrease in the severity of SMA. Type I patients generally have 1-2 copies of the *SMN2* gene, whereas type III patients may have 4 or more copies.[48,49] It has been hypothesized that, although the *SMN2* gene only produces 20% of the full length protein produced by *SMN1*, having an increased number of *SMN2* gene copies leads to production of a sufficient amount of SMN protein to make up for the production lost from the deletion or mutation of *SMN1*. This hypothesis, suggested by clinical data, was further verified in mouse models using human *SMN2* as a means of alleviating the symptoms of SMA-like mice lacking copies of the murine *Smn* gene. In 2000, two studies

showed that *Smn-/-* knockout mice that have no functional mouse *Smn* gene could be rescued from the disease phenotype by the addition of 8 copies of the human *SMN2* gene.[50,51] It is clear that *SMN2*, because of its ability to produce full length protein and modify the disease phenotype, represents a potential target for therapeutic intervention.

1.5 Therapeutic Approaches to Treat Spinal Muscular Atrophy

Treatment of alternative splicing-related diseases continues to be addressed and developed using a number of methods. Currently, most clinical treatments for diseases such as SMA are palliative, focusing on increasing the comfort of patients and helping to alleviate the symptoms associated with the disease.[52,53] Palliative treatments for SMA patients include physiotherapy to improve mobility and muscle strength, respiratory support to assist in routine breathing, and nutritional assistance via feeding tubes to circumvent difficulties in the consumption of food. For diseases such as cystic fibrosis, which have the potential for bacterial infections, routine antibiotic treatment is necessary to prevent additional complications. Although these treatments are beneficial to patients and can increase patient comfort, the underlying cause of the disease is often not addressed.

While there is currently no cure for SMA, a number of research groups have proposed treatments that address the underlying cause of the disease. These treatments target different processes associated with the transcription, splicing, and translation of the *SMN2* gene, so that more full length protein can be produced in the absence of *SMN1*. Although many have yet to reach the clinic, great potential has been shown for a number of different strategies, and further development will hopefully be able to make these strategies clinically feasible.

1.5.1 Small Molecule Therapeutics

Small molecule therapeutics for SMA have been examined in a number of contexts, from cellular assays to clinical trials, with moderate success (**Figure 1.4**). In 2001, Burghes and co-workers identified aclarubicin (**1.1**), an anthracycline antibiotic, which increases exon 7 inclusion in *SMN2* transcripts.[54] Aclarubicin has been shown to interact with DNA topoisomerase II, but how this increases exon 7 inclusion is unknown. The major disadvantage of this treatment is that it is non-specific and highly toxic, reducing the possibility of its development as a clinical treatment substantially. Sodium butyrate (**1.2**) and valproic acid (**1.3**), both histone deacetylase

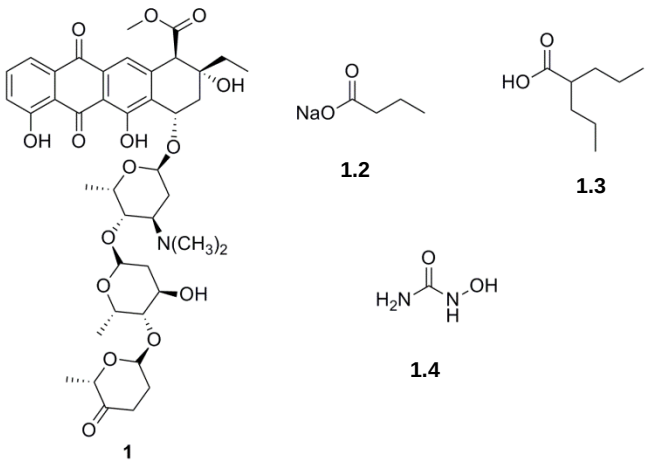


Figure 1.4. Small molecule treatments for SMA

inhibitors (HDACi), cause an increase in the level of SR protein transcript and *SMN2* transcript, respectively.[55,56] HDACi act by inhibiting the removal of acetyl groups from the lysine residues of histones, causing chromatin to remain unwound and able to interact with RNA polymerase, increasing expression of various genes. These molecules are non-specific, and exploration of the effects of using a non-specific transcription enhancer is ongoing. Several small molecules, including sodium butyrate, valproic acid, and hydroxyurea (**1.4**), are currently being utilized in human clinical trials.[57,58] These trials have only produced mild therapeutic effects in patients with less severe forms of the disease, indicating that further small molecule screening, as well as other strategies, must be utilized. Identification of novel compounds that affect *SMN2* transcription, splicing, and translation continues, and hopefully additional clinically relevant molecules will be identified.[59]

1.5.2. Gene Therapy

As SMA results from a lack of *SMN1* gene expression, several research groups have worked toward the development of a gene therapy method that would reconstitute the *SMN1* gene using a viral vector encoding the gene.[60] This method is one of the more direct means of replacing the absent SMN protein. However, possible issues associated with the use of a viral vector, as well as stable delivery and expression of the vector, have slowed its translation from the laboratory to the clinic. In mouse models, the vector has shown promise, helping to increase the levels of *SMN1* transcript and SMN protein as well as showing the best life expectancy outcome to date, suggesting that it is a viable method that should continue to be explored, despite possible issues associated with human use.[61] Further work on this treatment is ongoing.

1.5.3 Antisense Oligonucleotides

Antisense oligonucleotides have been widely utilized in a number of contexts, including the down-regulation and silencing of gene expression using RNA interference (RNAi) and to alter the splicing of RNA by blocking of splicing sequences in diseases like cystic fibrosis.[62] In the context of SMA, it was initially unknown if this strategy would be amenable to development as a therapy, as the goal of the antisense oligonucleotide was not to down-regulate or exclude a particular sequence element, but rather to include a sequence element. Several research groups explored the sequences within and surrounding exon 7 of *SMN2* in order to identify sequence elements that could be blocked in order to increase inclusion of exon 7 (**Figure 1.5**).

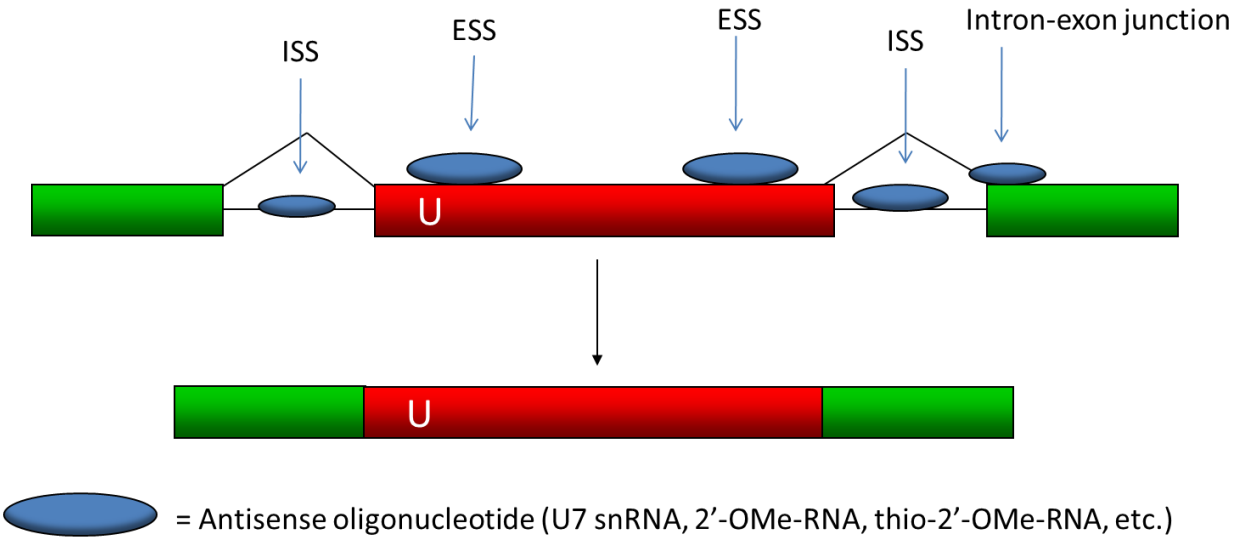


Figure 1.5. Possible targets for antisense oligonucleotides on *SMN2* pre-mRNA

In 2007, Krainer and co-workers conducted an antisense walk across exon 7 using antisense oligonucleotides of 15-16 bases to bind to different sequence elements of the exon and see what effect the oligonucleotides would have on splicing.[63] They identified a number of sequences that had positive effects (leading to exon 7 inclusion), as well as several sites that had the opposite effect, leading to exon 7 exclusion. These sites corresponded to the binding of other splicing proteins, including enhancers (such as SR-like protein Htra2β1) and inhibitors (hnRNP A1). Since that time, additional antisense walks have been conducted by Krainer and others, identifying additional sites on which antisense oligonucleotides can bind and increase inclusion of exon 7.[64,65] In general, these oligonucleotides block inhibitory factors such as hnRNPs from

binding *SMN2* pre-mRNA or redirect the splicing machinery away from exon 8 and toward exon 7. With their laboratory successes, antisense oligonucleotides have been further explored as possible clinically relevant molecules by commercial laboratories, and several are currently recruiting for phase I human clinical trials as potential treatments for SMA. Time will tell if they can continue to be a viable method for increasing exon 7 inclusion in *SMN2* and production of full length protein in humans, after showing great promise in cellular and mouse models.

1.5.4 Trans-Splicing

A novel strategy developed by Lorson and co-workers takes advantage of a normal but rarely used form of alternative splicing called "trans-splicing".[66,67] This process takes advantage of the multi-step process of mRNA splicing by allowing for the attachment of two different, non-linear mRNA pieces in the second step of the splicing reaction, instead of the normal attachment of the linear mRNA in two pieces separated by the intron. The resulting molecule is a hybrid of the two mRNAs, containing sequence elements from two different mRNA molecules. Lorson and co-workers recognized that this process can be enhanced if the second mRNA piece is connected to the target RNA branch point sequence via antisense oligonucleotide binding, increasing the effective concentration of the trans-splicing mRNA. Combining these traits, they developed a trans-splicing antisense oligonucleotide targeted to *SMN2* pre-mRNA via an RNA binding domain with a second region that contained the correct *SMN2* mRNA protein-encoding sequence (**Figure 1.6**). Attachment to and splicing of the trans-splicing antisense *SMN2* mRNA produced a trans-spliced mRNA product that was able to produce the correct SMN protein under cellular conditions, as measured by immunofluorescence.

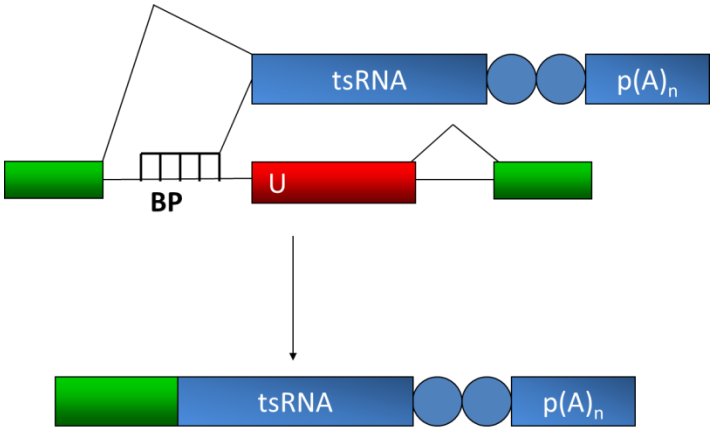


Figure 1.6. Trans-splicing mRNA strategy

Later work to further develop these molecules led to the incorporation of several additional features. First, an antisense walk across the *SMN2* gene indicated that exon 3 contained the most beneficial site for binding of their trans-splicing antisense oligonucleotide.[68]

Second, Lorson and co-workers found that inclusion of an antisense oligonucleotide targeted to the junction of exon 8 downstream of exon 7 increased the ability of the trans-splicing reaction to occur, leading to further SMN protein production.[69] Third, and most importantly, they were able to encode these molecules into viral vectors and stably express them in mouse models.[70,71] The vector contains both the trans-splicing oligonucleotide and the antisense oligonucleotide, creating a powerful method for altering mRNA splicing and increasing production of SMN protein. As mentioned, these vectors have been successful in mouse models, and it is only a matter of time before they end up in clinical trials.

1.5.5 Bifunctional Molecules

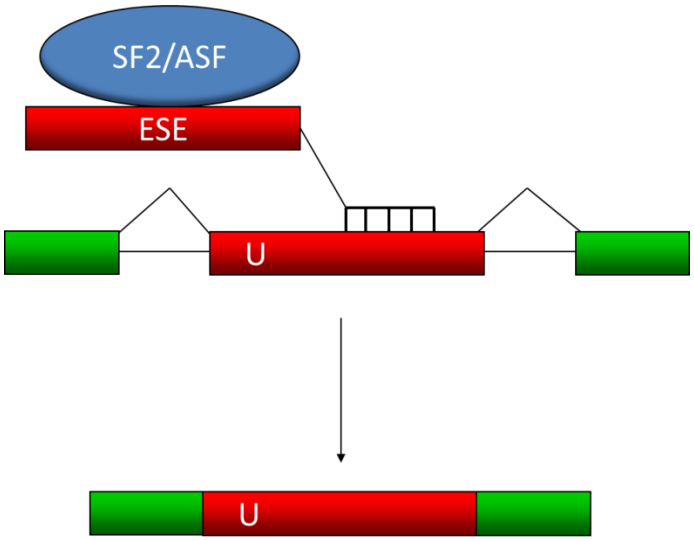


Figure 1.7. Bifunctional mRNA strategy

Antisense oligonucleotides can also be used in novel ways to mimic proteins or RNA binding sites, leading to changes in alternative splicing patterns. As with the oligonucleotides mentioned above, the goal of these strategies is to increase inclusion of an exon, but in this case the method is to reconstitute some binding event or molecular interaction that is lost under disease conditions. Two particular strategies have been used to increase inclusion of *SMN2* exon 7 and treat SMA. Mutoni and co-workers developed a bifunctional antisense oligonucleotide containing a sequence targeted to *SMN2* pre-mRNA and a second sequence containing multiple repeats of the exon 7 ESE sequence motif that is interrupted by the C→T transition in *SMN2* pre-mRNA (**Figure 1.7**).[72-74] The binding of this bifunctional oligonucleotide to the target pre-mRNA provides a location for SF2/ASF to bind, assisting in the recruitment of the remaining spliceosomal proteins as described previously. Lorson and co-workers were able to increase the inclusion of exon 7 and expression of the full length SMN protein under cellular conditions, and this method was amenable to encoding the

antisense oligonucleotide into a viral vector, making it a possible therapeutic option for SMA patients.[72]

A second novel strategy utilizing antisense oligonucleotides by Krainer and co-workers involved the development of SR protein mimics that could bind to the *SMN2* pre-mRNA and recruit the spliceosomal machinery via interaction with synthetic RS domains.[20] These molecules consisted of peptide nucleic acid (PNA) RNA binding domains and peptide regions consisting of 10 RS, RD, or RE dipeptide repeats. These molecules were all effective at increasing inclusion of exon 7 in *SMN2* mRNA, as well as *BRCA* mRNA. Further work by Caputi and co-workers demonstrated that a minimal RS domain of 8 dipeptide repeats was able to modulate the splicing of *Bcl-x* mRNA.[26] *Bcl-x* is a gene that encodes two proteins, Bcl-xL and Bcl-xS, that both act as regulators of cancer cell proliferation. They demonstrated that these molecules were effective at switching the splicing isoform from Bcl-xL, an anti-apoptotic form, to Bcl-xS, the pro-apoptotic form both *in vitro* and, importantly, *in vivo*. In both studies, a small number of minimal RS domains were utilized, but little was done to explore the structural space occupied by the synthetic RS domain and see if changes in the structure can affect activity. SR protein RS domains are intrinsically disordered, suggesting that addition of structural elements may change activity, but little work had been done previously to explore this area.

1.6 Project Design

The goal of this project is to develop specific synthetic alternative splicing activators that will be used to increase the knowledge of the splicing reaction and could ultimately be used as lead molecules for molecular therapeutics. These molecules would be most useful if they could be readily redesigned to address different types of splicing-related diseases. Taking this into consideration, our current design of alternative splicing activators focuses on bifunctional molecules containing a portion that specifically binds to the pre-mRNA and a portion that activates splicing. The overall approach that are using is shown in **Figure 1.8**.

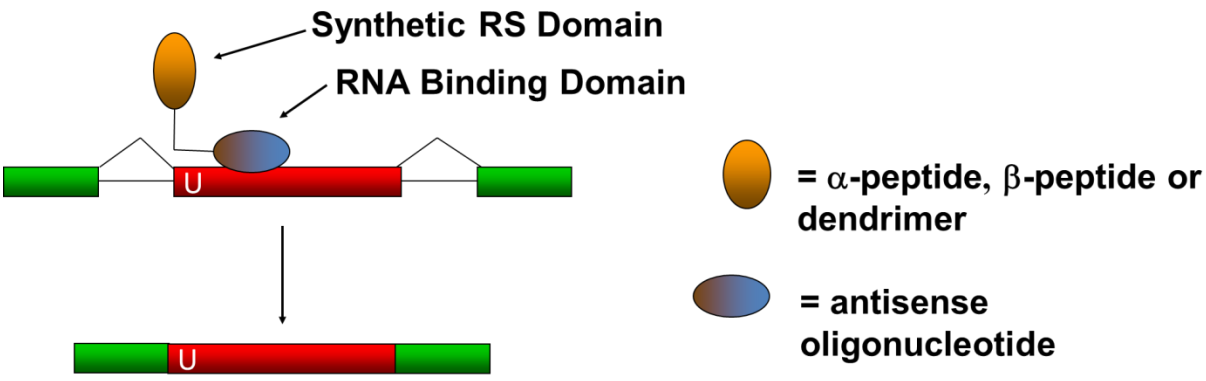


Figure 1.8. General approach for splicing activation by synthetic activators

We are focused in this project only on the understanding and development of the splicing activation domain of the bifunctional molecules, not the RNA binding domain. Therefore, we have used previously developed antisense oligonucleotides to specifically target the splicing activator to the *SMN2* pre-mRNA. The activation domain is modeled on a class of splicing activators, called the SR proteins, which contain phosphorylated RS dipeptide repeat sequences. These highly charged molecules recruit a collection of proteins (the spliceosome) to the pre-mRNA for correct splicing. We are exploring the sequence, charge, and structural requirements of the splicing activator using α-peptide, β-peptide, and dendrimer scaffolds. The efficacy of these molecules is examined using an *in vitro* splicing assay.

1.7 Importance and Relevance

Alternative splicing is an important process to understand, and the control of splicing has great potential as a therapeutic strategy. Spinal muscular atrophy is a common genetic disease that is currently without effective therapeutic treatment, but which could be treated by molecules that modulate alternative splicing of *SMN2* pre-mRNA. Increased levels of full-length *SMN2* mRNA containing exon 7 have been shown to ameliorate the disease in mouse models, and an increased copy number of *SMN2* in humans has been inversely correlated with disease severity, indicating that *SMN2* is an important target for therapeutic intervention. There has been significant exploration of small molecules to treat SMA, including molecules in clinical trials, but these treatments suffer from low efficacy and lack of specificity. Genetic approaches, such as antisense oligonucleotides and gene therapy, are promising, but are limited by delivery to and stability in the cellular environment. Minimal synthetic splicing activators could act specifically,

similar to other genetic approaches, and mimic the SR proteins that are unable to bind to *SMN2* pre-mRNA. The scope of possible RS domain mimics has not been explored in great detail previously, and the proposed synthetic activators introduce variations in sequence, structure, charge, and scaffold, representing the first time that non-peptide conjugates have been explored as splicing activators. These molecules will help to increase the knowledge of the requirements of minimal RS domains and provide a platform for the development of splicing activators that have strong potential to be therapeutic lead molecules.

1.8 References

(1) Johnson, J. M.; Castle, J.; Garrett-Engele, P.; Kan, Z.; Loerch, P. M.; Armour, C. D.; Santos, R.; Schadt, E. E.; Stoughton, R.; Shoemaker, D. D. *Science* **2003**, *302*, 2141.

(2) Pan, Q.; Shai, O.; Lee, L. J.; Frey, B. J.; Blencowe, B. J. *Nat. Genet.* **2008**, *40*, 1413-5.

(3) Wang, E. T.; Sandberg, R.; Luo, S.; Khrebtukova, I.; Zhang, L.; Mayr, C.; Kingsmore, S. F.; Schroth, G. P.; Burge, C. B. *Nature* **2008**, *456*, 470-6.

(4) Liu, H. X.; Chew, S. L.; Cartegni, L.; Zhang, M. Q.; Krainer, A. R. *Mol. Cell. Biol.* **2000**, *20*, 1063.

(5) Smith, C. W.; Valcarcel, J. *Trends Biochem. Sci.* **2000**, *25*, 381.

(6) Graveley, B. R. *RNA* **2000**, *6*, 1197.

(7) Manley, J. L.; Tacke, R. *Genes Dev.* **1996**, *10*, 1569.

(8) Sanford, J. R.; Ellis, J.; Caceres, J. F. *Biochem. Soc. Trans.* **2005**, *33*, 443.

(9) Chandler, S. D.; Mayeda, A.; Yeakley, J. M.; Krainer, A. R.; Fu, X. D. *Proc. Nat. Acad. Sci. U. S. A.* **1997**, *94*, 3596.

(10) Kashima, T.; Rao, N.; David, C. J.; Manley, J. L. *Hum. Mol. Genet.* **2007**, *16*, 3149.

(11) Hofmann, Y.; Wirth, B. *Hum. Mol. Genet.* **2002**, *11*, 2037.

(12) Haynes, C.; Iakoucheva, L. M. *Nucleic Acids Res.* **2006**, *34*, 305.

(13) Shen, H.; Green, M. R. *Mol. Cell.* **2004**, *16*, 363.

(14) Shen, H.; Kan, J. L.; Green, M. R. *Mol. Cell.* **2004**, *13*, 367.

(15) Singh, R.; Valcarcel, J. *Nat. Struct. Mol. Biol.* **2005**, *12*, 645.

(16) Fu, X. D. *RNA* **1995**, *1*, 663.

(17) Graveley, B. R.; Hertel, K. J.; Maniatis, T. *EMBO J.* **1998**, *17*, 6747.

(18) Allemand, E.; Gattoni, R.; Bourbon, H. M.; Stevenin, J.; Caceres, J. F.; Soret, J.; Tazi, J. *Mol. Cell. Biol.* **2001**, *21*, 1345.

(19) Caceres, J. F.; Krainer, A. R. *EMBO J.* **1993**, *12*, 4715.

(20) Cartegni, L.; Krainer, A. R. *Nat. Struct. Biol.* **2003**, *10*, 120.

(21) Stamm, S. *J. Biol. Chem.* **2008**, *283*, 1223.

(22) Nikolakaki, E.; Drosou, V.; Sanidas, I.; Peidis, P.; Papamarcaki, T.; Iakoucheva, L. M.; Giannakouros, T. *Biochim. Biophys. Acta* **2008**, *1780*, 214.

(23) Xiao, S. H.; Manley, J. L. *EMBO J.* **1998**, *17*, 6359.

(24) Wu, J. Y.; Maniatis, T. *Cell* **1993**, *75*, 1061.

(25) Lai, M. C.; Lin, R. I.; Tarn, W. Y. *Proc. Nat. Acad. Sci. U. S. A.* **2001**, *98*, 10154.

(26) Wilusz, J. E.; Devanney, S. C.; Caputi, M. *Nucleic Acids Res.* **2005**, *33*, 6547.

(27) Cartegni, L.; Chew, S. L.; Krainer, A. R. *Nat. Rev. Genet.* **2002**, *3*, 285.

(28) Garcia-Blanco, M. A.; Baraniak, A. P.; Lasda, E. L. *Nat. Biotechnol.* **2004**, *22*, 535.

(29) Wang, G. S.; Cooper, T. A. *Nat. Rev. Genet.* **2007**, *8*, 749.

(30) Cooper, T. A.; Wan, L.; Dreyfuss, G. *Cell* **2009**, *136*, 777.

(31) Poulos, M. G.; Batra, R.; Charizanis, K.; Swanson, M. S. *Cold Spring Harbor Perspect. Biol.* **2011**, *3*, a000778.

(32) Buratti, E.; Chivers, M.; Kralovicova, J.; Romano, M.; Baralle, M.; Krainer, A. R.; Vorechovsky, I. *Nucleic Acids Res.* **2007**, *35*, 4250.

(33) Roca, X.; Olson, A. J.; Rao, A. R.; Enerly, E.; Kristensen, V. N.; Borresen-Dale, A. L.; Andresen, B. S.; Krainer, A. R.; Sachidanandam, R. *Genome Res.* **2008**, *18*, 77.

(34) Gebski, B. L.; Mann, C. J.; Fletcher, S.; Wilton, S. D. *Hum. Mol. Genet.* **2003**, *12*, 1801.

(35) Goemans, N. M.; Tulinius, M.; van den Akker, J. T.; Burm, B. E.; Ekhart, P. F.; Heuvelmans, N.; Holling, T.; Janson, A. A.; Platenburg, G. J.; Sipkens, J. A.; Sitsen, J. M.; Aartsma-Rus, A.; van Ommen, G. J.; Buyse, G.; Darin, N.; Verschuuren, J. J.; Campion, G. V.; de Kimpe, S. J.; van Deutekom, J. C. *N. Engl. J. Med.* **2011**, *364*, 1513.

(36) Arambula, J. F.; Ramisetty, S. R.; Baranger, A. M.; Zimmerman, S. C. *Proc. Nat. Acad. Sci. U. S. A.* **2009**, *106*, 16068.

(37) Wong, C. H.; Fu, Y.; Ramisetty, S. R.; Baranger, A. M.; Zimmerman, S. C. *Nucleic Acids Res.* **2011**, *39*, 8881.

(38) Munsat, T. L.; Davies, K. E. *Neuromuscul. Disord.* **1992**, *2*, 423.

(39) Lefebvre, S.; Burglen, L.; Reboullet, S.; Clermont, O.; Burlet, P.; Viollet, L.; Benichou, B.; Cruaud, C.; Millasseau, P.; Zeviani, M.; et al. *Cell* **1995**, *80*, 155.

(40) Eggert, C.; Chari, A.; Laggerbauer, B.; Fischer, U. *Trends Mol. Med.* **2006**, *12*, 113.

(41) Meister, G.; Eggert, C.; Fischer, U. *Trends Cell. Biol.* **2002**, *12*, 472.

(42) Lorson, C. L.; Hahnen, E.; Androphy, E. J.; Wirth, B. *Proc. Nat. Acad. Sci. U. S. A.* **1999**, *96*, 6307.

(43) Monani, U. R.; Lorson, C. L.; Parsons, D. W.; Prior, T. W.; Androphy, E. J.; Burghes, A. H.; McPherson, J. D. *Hum. Mol. Genet.* **1999**, *8*, 1177.

(44) Setola, V.; Terao, M.; Locatelli, D.; Bassanini, S.; Garattini, E.; Battaglia, G. *Proc. Nat. Acad. Sci. U. S. A.* **2007**, *104*, 1959.

(45) Cartegni, L.; Krainer, A. R. *Nat. Genet.* **2002**, *30*, 377.

(46) Kashima, T.; Manley, J. L. *Nat. Genet.* **2003**, *34*, 460.

(47) Cartegni, L.; Hastings, M. L.; Calarco, J. A.; de Stanchina, E.; Krainer, A. R. *Am. J. Hum. Genet.* **2006**, *78*, 63.

(48) Brahe, C. *Neuromuscul. Disord.* **2000**, *10*, 274.

(49) Feldkotter, M.; Schwarzer, V.; Wirth, R.; Wienker, T. F.; Wirth, B. *Am. J. Hum. Genet.* **2002**, *70*, 358.

(50) Hsieh-Li, H. M.; Chang, J. G.; Jong, Y. J.; Wu, M. H.; Wang, N. M.; Tsai, C. H.; Li, H. *Nat. Genet.* **2000**, *24*, 66.

(51) Monani, U. R.; Sendtner, M.; Coovert, D. D.; Parsons, D. W.; Andreassi, C.; Le, T. T.; Jablonka, S.; Schrank, B.; Rossoll, W.; Prior, T. W.; Morris, G. E.; Burghes, A. H. *Hum. Mol. Genet.* **2000**, *9*, 333.

(52) Mercuri, E.; Bertini, E.; Iannaccone, S. T. *Lancet Neurol.* **2012**, *11*, 443.

(53) Wang, C. H.; Finkel, R. S.; Bertini, E. S.; Schroth, M.; Simonds, A.; Wong, B.; Aloysius, A.; Morrison, L.; Main, M.; Crawford, T. O.; Trela, A.; Participants of the International Conference on, S. M. A. S. o. C. *J. Child Neurol.* **2007**, *22*, 1027.

(54) Andreassi, C.; Jarecki, J.; Zhou, J.; Coovert, D. D.; Monani, U. R.; Chen, X.; Whitney, M.; Pollok, B.; Zhang, M.; Androphy, E.; Burghes, A. H. *Hum. Mol. Genet.* **2001**, *10*, 2841.

(55) Brichta, L.; Hofmann, Y.; Hahnen, E.; Siebzehnrubl, F. A.; Raschke, H.; Blumcke, I.; Eyupoglu, I. Y.; Wirth, B. *Hum. Mol. Genet.* **2003**, *12*, 2481.

(56) Chang, J. G.; Hsieh-Li, H. M.; Jong, Y. J.; Wang, N. M.; Tsai, C. H.; Li, H. *Proc. Nat. Acad. Sci. U. S. A.* **2001**, *98*, 9808.

(57) Swoboda, K. J.; Kissel, J. T.; Crawford, T. O.; Bromberg, M. B.; Acsadi, G.; D'Anjou, G.; Krosschell, K. J.; Reyna, S. P.; Schroth, M. K.; Scott, C. B.; Simard, L. R. *J. Child Neurol.* **2007**, *22*, 957.

(58) Tsai, L. K.; Yang, C. C.; Hwu, W. L.; Li, H. *Eur. J. Neurol.* **2007**, *14*, e8.

(59) Xiao, J.; Marugan, J. J.; Zheng, W.; Titus, S.; Southall, N.; Cherry, J. J.; Evans, M.; Androphy, E. J.; Austin, C. P. *J. Med. Chem. 54*, 6215.

(60) Passini, M. A.; Cheng, S. H. *Trends Mol. Med.* **2011**, *17*, 259.

(61) Glascock, J. J.; Osman, E. Y.; Wetz, M. J.; Krogman, M. M.; Shababi, M.; Lorson, C. L. *Hum. Gene. Ther. 23*, 330.

(62) Kole, R.; Krainer, A. R.; Altman, S. *Nat. Rev. Drug Discovery* **2012**, *11*, 125.

(63) Hua, Y.; Vickers, T. A.; Baker, B. F.; Bennett, C. F.; Krainer, A. R. *PLoS Biol.* **2007**, *5*, e73.

(64) Hua, Y.; Sahashi, K.; Hung, G.; Rigo, F.; Passini, M. A.; Bennett, C. F.; Krainer, A. R. *Genes Dev. 24*, 1634.

(65) Hua, Y.; Vickers, T. A.; Okunola, H. L.; Bennett, C. F.; Krainer, A. R. *Am. J. Hum. Genet.* **2008**, *82*, 834.

(66) Garcia-Blanco, M. A. *J. Clin. Invest.* **2003**, *112*, 474.

(67) Coady, T. H.; Shababi, M.; Tullis, G. E.; Lorson, C. L. *Mol. Ther.* **2007**, *15*, 1471.

(68) Shababi, M.; Lorson, C. L. *J. Mol. Neurosci.* **2012**, *46*, 459.

(69) Coady, T. H.; Baughan, T. D.; Shababi, M.; Passini, M. A.; Lorson, C. L. *PLoS One* **2008**, *3*, e3468.

(70) Shababi, M.; Glascock, J.; Lorson, C. L. *Hum. Gene. Ther. 22*, 135.

(71) Coady, T. H.; Lorson, C. L. *J. Neurosci.* **2010**, *30*, 126.

(72) Baughan, T.; Shababi, M.; Coady, T. H.; Dickson, A. M.; Tullis, G. E.; Lorson, C. L. *Mol. Ther.* **2006**, *14*, 54.

(73) Marquis, J.; Meyer, K.; Angehrn, L.; Kampfer, S. S.; Rothen-Rutishauser, B.; Schumperli, D. *Mol. Ther.* **2007**, *15*, 1479.

(74) Skordis, L. A.; Dunckley, M. G.; Yue, B.; Eperon, I. C.; Muntoni, F. *Proc. Nat. Acad. Sci. U. S. A.* **2003**, *100*, 4114.

CHAPTER 2

DESIGN AND SYNTHESIS OF PEPTIDE SR PROTEIN MIMICS

2.1 Introduction

As the structure and function of proteins have been elucidated by the use of different biological and analytical techniques, it has become possible to determine which structures are important for activity and to develop molecules that are able to reconstitute the active site without the remaining protein structure. Development of these novel protein mimics has produced molecules that are active in a number of different contexts, including but not limited to binding of activator sites to activate protein function, inhibition of protein-protein interactions, and reconstitution of protein activity using a peptide mimic.[1-5] Thus, the development of novel molecules that are able to act as protein mimics is an important field that demands attention.

2.1.1 SR Protein Mimic Design

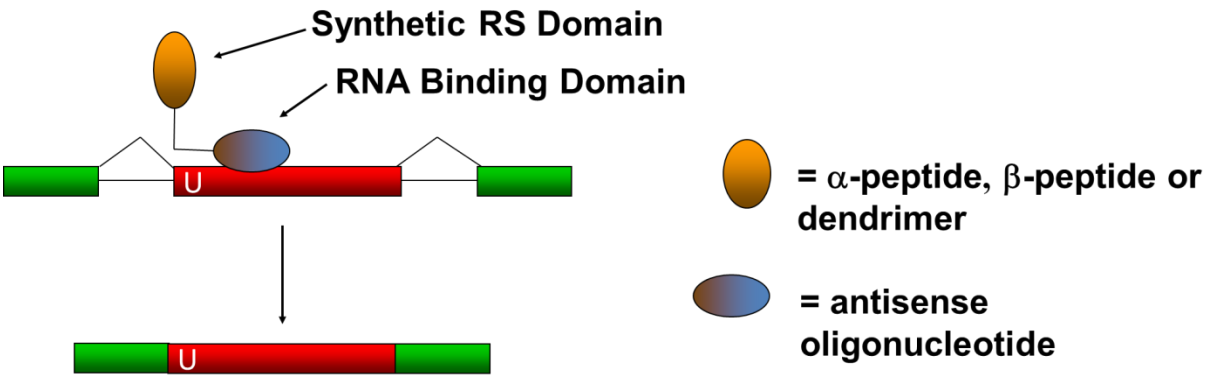


Figure 2.1: SR Protein Mimic Design

Our protein mimic design aims to recapitulate the activity of the SR protein SF2/ASF, a splicing factor involved in recognition of exon 7 in *SMN2* pre-mRNA that has reduced affinity due to the change in the ESE sequence. Taking inspiration from Krainer and co-workers,[5] the designed protein mimics are peptide-oligonucleotide conjugates containing two regions: first, an antisense oligonucleotide that can replace the RBD by interacting with a sequence on the *SMN2* pre-mRNA; and second, a synthetic RS domain that will mimic the natural RS domain and interact with other proteins via electrostatic interactions (**Figure 2.1**). The antisense oligonucleotide consists of a single-stranded DNA molecule, chosen because it should be relatively stable to nuclease degradation and is easily and cheaply synthesized.

While the focus of the research was on the synthetic RS domain, there is a great deal of potential for study of the antisense oligonucleotide. Work conducted by a number of research groups has shown that there is great potential for the use of antisense oligonucleotides to control alternative splicing outcomes (see Chapter 1).[6-10] We chose to target nucleotides +7-18 on exon 7, as they have been shown to be a good target for splicing enhancement and have been targeted by Krainer and co-workers in the development of their SR protein mimics.[5,7]

2.1.2 RS Domain Mimics – α-Peptide

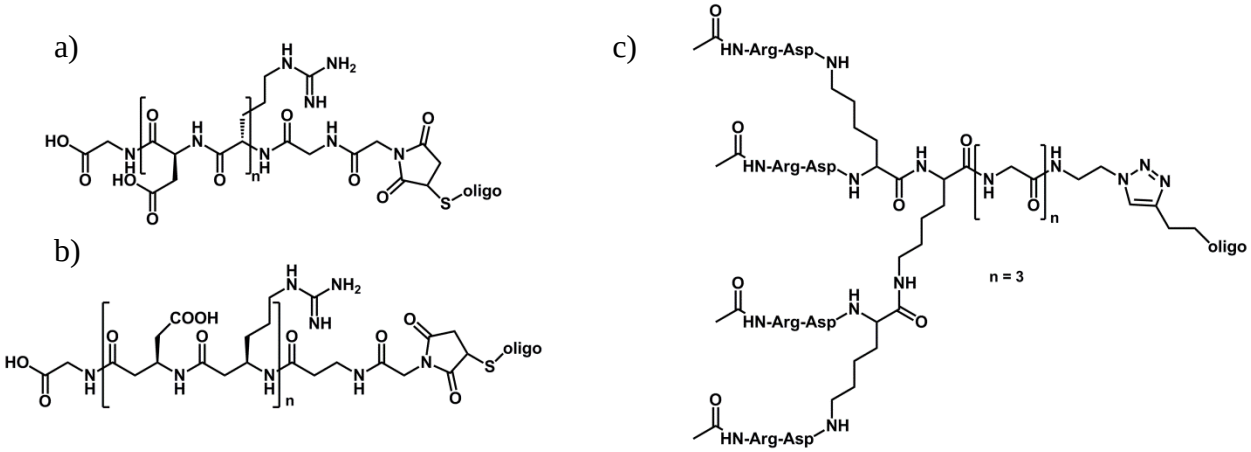


Figure 2.2: Synthetic RS Domain General Design. a) α-Peptide. b) β-Peptide. c) Peptide Dendrimer

The synthetic RS domain of the SR protein mimics consists of three motifs – α-peptide, β-peptide, or peptide dendrimer (**Figure 2.2**). α-Peptides have been used previously as RS domain mimics. We have chosen β-peptides as non-natural, linear α-peptide mimics that are protease resistant and able to form secondary structures. Peptide dendrimers were chosen for their protease stability and globular arrangement of amino acid residues. The goal of our research was to explore the effectiveness of these three motifs, determining which of the three has the best activity, as well as to compare different properties within these motifs. By using short peptides, we would be able to systematically analyze a single property, such as charge, without the potential interference or interactions of a larger protein.

2.1.2.1 Varying Dipeptide Repeats

Previous work has shown that synthetic RS domains consisting of 10 RS, RD, and RE dipeptide repeats are effective at changing splicing patterns in *BRCA1*, *SMN*, and *Bcl-x* pre-

mRNA. We have chosen to use the RD dipeptide in order to mimic the hyperphosphorylated state of the RS domain that is thought to be important for mediating these interactions. Phosphorylation of the RS domain plays an important role in mediating protein-protein interactions, preventing non-specific binding, and facilitating nuclear transport.[11-13] Although there may be some concern about mimicking the hyperphosphorylated state for nuclear transport, as dephosphorylation of the RS domain is necessary to transport the SR proteins out of the nucleus, the use of an *in vitro* splicing assay can avoid any potential issues associated with traversing the cellular membrane and nuclear transport. The arginine present in the synthetic RS domain may also facilitate transport into the cell, should these molecules be tested under *in vivo* conditions. Arginine-based cell-penetrating peptides (CPPs) have been used as carriers for antisense oligonucleotides to cross the cell membrane.[14] Caputi and co-workers demonstrated some activity using synthetic SR protein mimics in HeLa cells. Previous synthetic RS domains contained 8-10 dipeptide repeats, all of which had activity under *in vitro* conditions, but smaller iterations of the RS domain have not been explored. We chose to analyze minimal activators of 5, 8, and 10 dipeptide repeats to determine if they are sufficient to alter the splicing pattern of *SMN2* pre-mRNA.

2.1.2.2 Varying Charged Residues

The RS domain of SR proteins is highly charged, with positively charged arginine residues and serine residues that are negatively charged when phosphorylated. It has been shown that the highly charged nature of the RS domain is a necessary characteristic for facilitating protein-protein interactions. Although the protein may still have activity in the absence of these highly charged residues, less is known about the effects of charge removal in peptide models. We hypothesized that removal of either arginine or aspartic acid from our peptide mimics would decrease splicing activation, and thus designed several peptides in which alanine is substituted for either member of the dipeptide. In addition, peptides in which one or several of the dipeptides are replaced with alanine were designed in order to examine how removal of some, but not all, of the charged residues would affect splicing activation. Examination of SR proteins indicates that the presence of RSRS tetrapeptides is important for activity, so substitution of alanine residues while still maintaining RDRD tetrapeptides on the periphery of the synthetic RS domain can address whether this property holds true in a peptide model.

2.1.2.3 Varying Charge Distribution

To the best of our knowledge, the importance of sequence in the RS domain has not been explored extensively in peptide models. In particular, little work has been done to examine whether the RS dipeptides need to be arranged in alternating order to maintain activity, or if other arrangements still produce active molecules. We have designed several motifs in which the arrangement of arginine and aspartic acid residues is changed to an RRDD motif, with one or several repeats present on the peptide. We hypothesize that the arrangement of these residues will not have a large effect on activity, as the electrostatic interactions of the charged residues play the biggest role in protein-protein interactions, and the RS domains of the SR proteins are intrinsically disordered. These redistributed peptides will provide insight into how the charged residues interact with their protein targets, and whether or not the alternating charges are necessary for activity.

2.1.3 RS Domain Mimics – β-Peptide

Although the RS domains belong to a class of intrinsically disordered proteins, there is evidence based on characterization of small RD and RS repeat peptides that these sequences may be at least transiently structured.[15-17] This observed sequence tolerance suggest that there may be other structural solutions to splicing activation. With this in mind, we have designed a series of peptide-like molecules as substitutes for the α-peptide motif. β-Peptides are a useful alternative due to their protease resistance and similarity to α-peptides in their linear arrangement of amino acid residues.

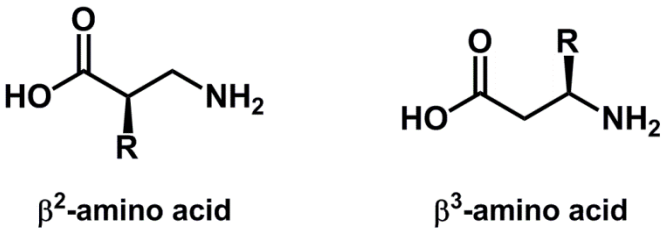


Figure 2.3: β-Amino Acids

β-Peptides are composed of β-amino acids, which are essentially α-amino acids containing an extra methylene carbon with side chain functionality at the α-position (β^2-amino acids) or the β-position (β^3-amino acids) (**Figure 2.3**). This modification of the natural amino

acid structure confers protease resistance and conformational flexibility. β-Peptides can fold into a variety of sequence-dependent secondary structures, such as 12- and 14-helices, in as few as four amino acids.[2] β-peptides have been used in a variety of biological applications, most importantly in the disruption of protein-protein interactions. In 2004, Schepartz and co-workers designed a series of 14-helices that were able to inhibit the formation of the hDM2-p53 complex, an interaction important in cancer proliferation.[1] Helix formation was facilitated by hydrophobic and charged residues that would favorably interact only in a helical structure, even under aqueous conditions.[18] β-peptide interaction with mRNA was demonstrated by Gellman and co-workers, who designed a β-peptide mimic of the tat protein that interacted with TAR RNA, an important interaction for HIV disease progression.[19] These studies demonstrated that β-peptides can be rationally designed to incorporate structure, and that they can interact with biologically relevant molecules.

2.1.3.1 Varying Dipeptide Repeats

The overall design of the β-peptides is similar to that of the α-peptides. A synthetic RS domain is linked to an oligonucleotide that is complementary to the *SMN2* pre-mRNA. The first set of β-peptides contain the same sequences of 5, 8, and 10 dipeptide repeats as the α-peptides, so the molecules containing the same number of residues can be compared directly. This can help to determine if a synthetic RS domain mimic with protease resistance has similar activity to molecules that have already been used in other studies.

2.1.3.2 Effects of Carbon Backbone

An interesting property of the β-peptides is the presence of the additional methylene carbon in the backbone of the β-amino acids, which increases protease resistance and conformational flexibility. The presence of this additional carbon presents the possibility of exploring the importance of the carbon backbone on activity of synthetic RS domain mimics, a property that has not been explored previously. We have designed a series of β-peptides that have a similar number of carbons as the backbones of α-peptides. As each β-amino acid has an additional carbon in the backbone, these β-peptides use fewer amino acids than their α-peptide counterparts (**Table 2.1**). Comparing these α- and β-peptides will suggest the effect the length of the backbone has on activity.

Table 2.1. Comparison of α- and β- peptides with similar backbone lengths

Peptide Type	# DR Repeats	Carbon Chain Length
α	5	34
β	4	36
α	8	57
β	6	52
α	10	63
β	7	60

2.1.3.3 Secondary Structure

The β-peptides have a unique ability to form secondary structures in as few as four amino acids through favorable interactions of the β-amino acid side chains. This allows for the possibility of "encoding" secondary structure into the β-peptide. We have designed a series of molecules that have the ability to form secondary structures in order to probe the ability of these polymers to alter splicing patterns. Helices will contain one to three faces with alternating R and D residues, which will serve the dual purpose of stabilizing the helix and providing the same array of positive and negative charges inherent to the RS domain.

2.2 Results and Discussion

2.2.1 Peptide RS Domain Synthesis

Both α- and β-peptide were synthesized on 2-chlorotrityl resin functionalized with glycine using standard Fmoc/OtBu solid-phase peptide synthesis (SPPS) procedures. This resin was chosen because the growing peptides could be cleaved under weakly acidic conditions using hexafluoroisopropanol (HFIP) without side-chain deprotection, allowing for monitoring via electrospray ionization (ESI) or matrix assisted laser desorption/ionization (MALDI) mass spectrometry (MS). Peptide synthesis was also monitored by the ninhydrin test developed by Kaiser and co-workers that utilizes the color-changing properties of ninhydrin in the presence of free amines. Purification of the peptides was done using reverse-phase high performance liquid chromatography (RP-HPLC) after cleavage from the resin using a TFA cleavage mix (see Materials and Methods). The peptides were cleaved, deprotected, precipitated with diethyl ether, redissolved in a mixture 1:1 mixture of H_2O:CH_3CN containing 0.1% TFA, filtered, and

analyzed by RP-HPLC. Purification of the peptides by RP-HPLC was not trivial, as all peptides had slightly different retention times due to changes in the charge profile (by substitution of charged residues for alanine) or size. Peptides mass was verified by ESI-MS or MALDI-MS before being used in further manipulations.

2.2.1.1 α-Peptide Synthesis

Table 2 lists the different peptides synthesized by SPPS. All α-peptides were synthesized with an N-terminal glycine residue (for flexibility at the peptide-oligonucleotide linkage) and a linker molecule. α-Peptides containing 5, 8, and 10 RD dipeptide repeats were synthesized, purified by RP-HPLC, and analyzed by MALDI-MS. These peptides were designed to explore the minimal number of charges needed to produce an effective splicing activator. In addition, α-peptides containing 8 dipeptide repeats of either AD or RA were synthesized in order to analyze the importance of charged moieties. These molecules were somewhat more difficult to purify, as the overall molecules were highly charged (as opposed to the RD dipeptides, which had a low overall charge due to cancellation of positive and negative charges), requiring careful examination of purification conditions to isolate them properly. Peptides containing one or two DDRR repeats in between DR dipeptide repeats have also been synthesized and isolated as described above.

Table 2.2. Synthesized α- and β-peptides

α-Peptides	**β-Peptides**
$(DR)_5$	$(\beta D\text{-}\beta R)_4$
$(DR)_8$	$(\beta D\text{-}\beta R)_5$
$(DR)_{10}$	$(\beta D\text{-}\beta R)_6$
$(DA)_8$	$(\beta D\text{-}\beta R)_7$
$(AR)_8$	$(\beta D\text{-}\beta R)_8$
$(DR)_3A(DR)_3$	$(\beta D\text{-}\beta R)_{10}$
$(DR)_3DDRR(DR)_3$	
$(DR)_3(DDRR)_2(DR)_3$	

2.2.1.2 β-Peptide Synthesis

β-Peptide analogs of α-peptides containing 5, 8, and 10 RD dipeptide repeats have been synthesized. Synthesis was conducted using commercially available β^3-R and β^3-D amino acids. β^3-Alanine was substituted for glycine at the N-terminus, followed by the linker molecule. In addition, β-peptides were synthesized containing different numbers of dipeptide repeats that were designed to contain similar backbone chain lengths to α-peptides while containing fewer charged residues (**Table 2**). Products were analyzed by ESI-MS and MALDI-MS, and purified as described above. Unfortunately, few of these peptides were conjugated and analyzed for alternative splicing modulation activity. These peptides were synthesized well before we began our conjugation reactions, and had degraded before we could use them in our reactions. The β-peptides will not be mentioned in subsequent discussion of molecule activity (see Chapter 4).

2.2.1.3 Peptide Synthesis Troubleshooting

The synthesis of α- and β-peptides, while straightforward, was not without difficulties. Reagent purity was especially important. PyBOP and HOBt coupling reagents, for example, both degrade over time at room temperature, decreasing coupling efficiency. Both reagents have been stored at 4°C to prevent degradation. Additional coupling reagents were utilized during peptide synthesis with varying degrees of success. HBTU, a benzotriazole coupling reagent similar to PyBOP, showed similar levels of efficiency and could be used in the absence of HOBt, which has limited availability due to its recent classification as an explosive. COMU, a coupling reagent based on ethyl (hydroxyimino)cyanoacetate (Oxyma) developed by Sigma Aldrich, was chosen as an alternative to PyBOP due to its reported high coupling efficiency and strong inhibition of racemization as a coupling additive.[20] Like PyBOP and HBTU, this coupling reagent had reasonable efficiency when used on longer peptides, losing coupling activity after extended use. The decreased efficiency of all coupling reagents over time highlighted the importance of maintaining fresh reagents for coupling reactions. Like coupling reagents, the efficiency of the Fmoc-cleaving agent piperidine decreased over time, requiring regular replacement in order to make sure that the cleavage reaction remained quantitative.

Water can play an inhibitory role in amide bond formation, as it can act as a nucleophile and cause the activated ester to convert back to the free acid moiety. To prevent water from interfering with the coupling reaction, reagents were extensively dried by azeotropic distillation with dry DMF, and dry DMF stored over activated molecular sieves was used in all coupling reagents. It was found that the efficiency of the reaction decreased with the increasing number of amino acid residues, making the synthesis of 10 RD dipeptide repeat molecules particularly challenging. This was due in part to the aggregation of peptides through backbone interactions while on the resin. Pre-swelling of the resin using CH_2Cl_2, increasing reaction time, and conducting consecutive reactions with fresh coupling reagents helped to alleviate this problem.

Isolation of the peptides following cleavage and purification was not trivial. We were able to isolate the crude peptide with side-chain protecting groups removed, and precipitation with ethyl ether to remove additives from the TFA cleavage mix and cleaved side chains was moderately successful. There remained some side-products in the peptide mixture that we attempted to remove by RP-HPLC. Unfortunately, some product was lost during RP-HPLC purification, despite collection of the majority of the solvent producing a UV signal. It is likely that issues with the HPLC pump, particularly for the CH_3CN solvent, played a role in the signal broadening that prevented collection of the entire sample. It was unknown if this was due to the viscosity of CH_3CN or the pump itself, although it appeared to be functioning normally. We attempted to address these issues by dissolving the peptide cleavage mixture in organic solvents and filtering off remaining solids, which had a somewhat beneficial effect. Further experimentation with peptide cleavage mix solubility and solvent combinations would help with increasing the yield following RP-HPLC purification, improving the quality of the peptides being used for further manipulation.

2.2.2 Peptide-Oligonucleotide Conjugation

The conjugation of synthetic RS domains to antisense oligonucleotides is not trivial, and some thought must be put into deciding the conjugation method that will work most efficiently with the two fragments involved. In their design, Krainer and co-workers used PNA as their antisense oligonucleotide, meaning that the entire SR protein mimic could be synthesized in an iterative fashion without the need for a conjugation reaction. In our design, the antisense oligonucleotide and peptide are synthesized separately. Thus, we needed to choose an

orthogonal reaction scheme that would allow for conjugation of the two fragments with little potential for side-product formation.

The field of bioconjugation is extensive, and many strategies are available for connection of peptide and oligonucleotide fragments. Our criteria for choosing a particular method included low cost to purchase or synthesize the conjugates, that the coupling reaction be orthogonal to other possible reactive groups, and that the reaction conditions be mild to ensure stability of the peptide and oligonucleotide fragments. Commercially available linker molecules provide fairly straightforward synthesis and conjugation reactions, but the fairly high cost associated with their use drove us in a different direction. A number of possible conjugation schemes that would utilize commercially available oligonucleotide modifiers. Of those available, three were chosen: maleimide-thiol addition, Diels Alder cycloaddition, and copper-catalyzed alkyne azide cycloaddition (CuAAC or 'click chemistry') (**Figure 2.4**).

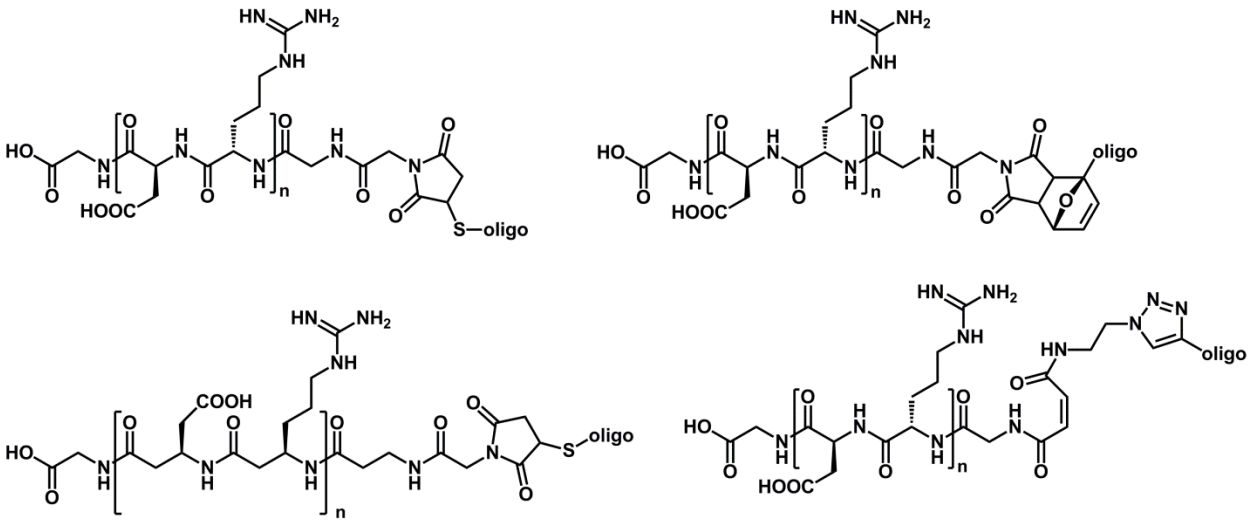
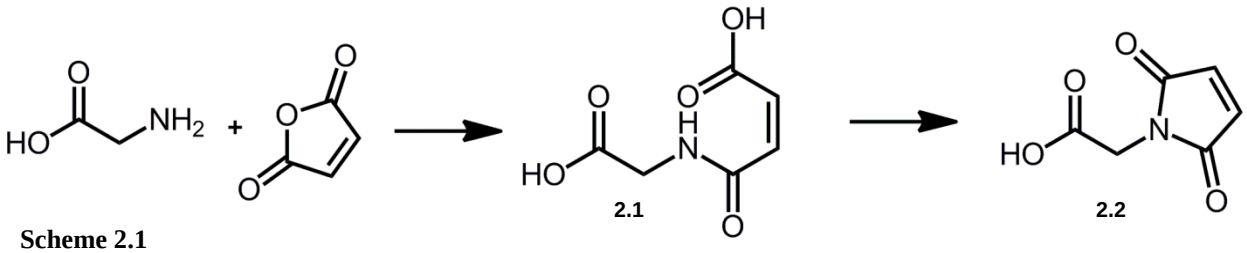


Figure 2.4. Possible Peptide-Oligonucleotide Conjugation Schemes

2.2.2.1 Maleimide-Thiol Addition

The maleimide-thiol addition has been used in a wide variety of contexts to produce conjugated molecules. It has been used to produce peptide-oligonucleotide conjugates using a number of biologically relevant starting materials. This reaction consists of two components – a maleimide ring containing an α,β-conjugated carbonyl system that acts as a Michael acceptor, and a free thiol that acts as a Michael donor. Under buffered conditions, the free thiol will

undergo nucleophilic attack of the alkene, forming an enol functionality. Protonation of the enol produces the ketone and the substituted final product. This reaction is advantageous in that it is orthogonal to the functionality of the peptide and oligonucleotide, can be run under mild, slightly basic buffer conditions, and the cost of starting materials is cheap. The maleimide linker can be easily synthesized from commercially available reagents in large quantities, and it is able to withstand TFA mediated cleavage after addition to the synthetic peptide RS domains during solid phase peptide synthesis. The thiol functionality on the oligonucleotide can be added at the 3'-end by means of a commercially available modifier (Glen Research), and the only cumbersome step is reduction of the disulfide to free the thiol functionality prior to coupling.



Scheme 2.1

Synthesis of the maleimide linker is a straightforward preparation from glycine and maleic anhydride (**Scheme 2.1**). In the first step, nucleophilic attack of the glycine nitrogen forms an amide bond, opening the electrophilic anhydride. Addition of base and refluxing the reaction in toluene allows for a second attack of the free lone pair of nitrogen on the second acid moiety, closing the maleimide and releasing water from the reaction. This series of reactions is straightforward and high yielding, producing a large quantity of the linker that can be used for coupling. As the molecule is built off of glycine, it can be directly linked to the peptide at the N-terminus as the last coupling step before the peptide is cleaved from the resin. The molecule is stable to the TFA-mediated cleavage from the resin, leaving it available for coupling to the oligonucleotide.

Scheme 2.2

The maleimide-thiol addition was examined under a number of conditions in order to determine what would be the most successful at producing peptide-oligonucleotide conjugates (**Scheme 2.2**). In 1994, Ede et al. coupled peptides and oligonucleotides in a buffer of 20% 0.1M triethylammonium acetate (TEAA) in CH_3CN, and this method was initially chosen to produce our conjugates.[21] The buffer system was chosen because it was the buffer system utilized in this report to purify the oligonucleotide and remove excess dithiothreitol (DTT) prior to conjugation. We initially chose DTT as the disulfide reducing agent, as it is a common reagent for reducing disulfide bonds. Initial attempts using these reaction conditions were moderately successful, producing a small amount of product relative to the starting reagents. Several factors were explored to improve the yield of these reactions. It was first noted that the coupling buffer should be degassed to reduce the chance of oxidation of the thiol back to a disulfide. Next, the amount of time needed to fully reduce the disulfide was explored, as the time of 1-2 hours listed in the protocol was insufficient. Examination of the oligonucleotide at different time points revealed that overnight reduction with DTT was necessary for complete reduction, and conjugation with this fully reduced oligonucleotide produced a high yield of conjugate. Reaction solvent was also of a concern. The reaction needed to be slightly basic to promote thiol deprotonation, but not so basic so as to cause opening of the maleimide ring, which is sensitive to base. Addition of potassium hydroxide to the previously established buffer conditions to increase pH was moderately successful, although it appeared to lead to more opening of the maleimide ring than the desired conjugate. Examination of the pH of our buffers indicated that a reaction pH of around 7.5 was sufficient for this reaction to occur without too much breakdown of the malemide ring.

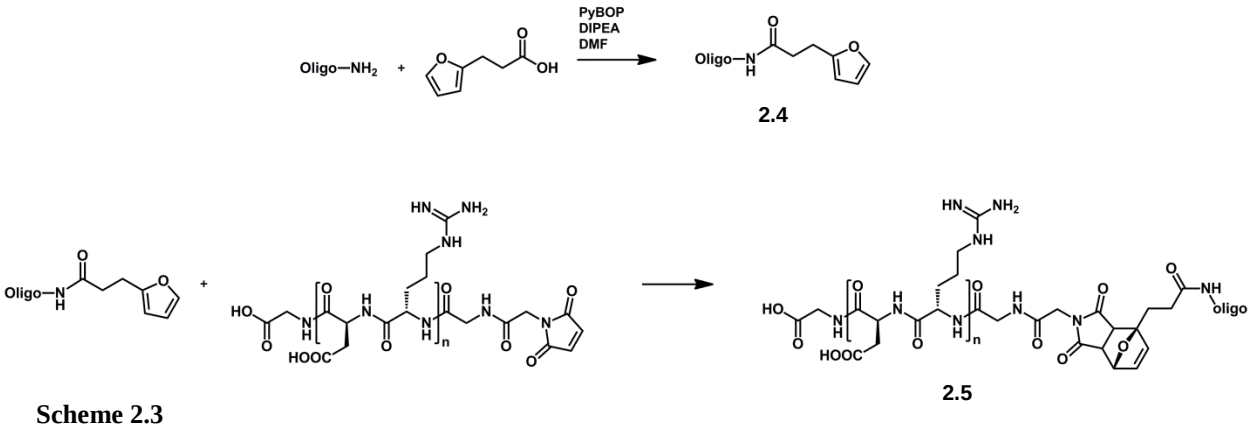
It was noted that an issue with performing RP-HPLC purification of the oligonucleotide was the substantial loss of the oligonucleotide after purification, reducing the overall yield. Alternative methods were sought to remove the DTT reducing agent, including precipitation with silver nitrate and extraction with organic solvents such as ethyl acetate. These methods, while moderately successful at removing DTT and ensuring that a majority of oligonucleotide remained available for coupling, were not 100% efficient at removing DTT. This reducing agent also contains free thiols, and the presence of excess DTT meant that the free maleimide would often react preferentially with the DTT rather than the thiol on the oligonucleotide. Thus, we explored alternatives to DTT as a reducing agent.

We determined that *tris*(2-carboxyethyl)phosphine (TCEP) would be a suitable alternative. TCEP is a phosphine reducing agent that is advantageous over DTT for several reasons: it is a more powerful reducing agent, an irreversible reducing agent, less susceptible to oxidation in air, and less reactive with malemide functionality, although there is some evidence that it can react with maleimides under certain conditions.[22] It can also be purchased on-resin, making it much easier to remove prior to conjugation. We tested TCEP both as a 0.5M solution and on resin, with both providing very good reduction of the disulfide bond as indicated by MALDI-MS. Importantly, reduction using TCEP could be completed in less than an hour, greatly improving the speed with which we could synthesize our conjugates. On resin, the TCEP was removed by centrifugation, whereas in solution it was left in the reaction mixture. In either case, we were able to obtain good yields of the maleimide-thiol conjugate, even in the presence of excess TCEP. The experimental section contains a list of assembled conjugates.

Purification of the maleimide-thiol conjugation utilized several methods, including purification by Amicon Ultra-0.5 mL Centrifugal Filters (Millipore) and RP-HPLC. The Amicon columns contain a 3000 Da molecular weight cutoff membrane that would allow the starting materials to pass through the membrane while retaining the conjugate, which had an average molecular weight around 6000 Da. This method was moderately successful, isolating primarily the desired product but occasionally retaining any excess oligonucleotide. This was likely because the molecular weight of the oligonucleotides was just slightly larger than the membrane cutoff. Although the manufacturer's notes indicate that it should pass through the membrane (recommending that the molecule be at least twice the size of the membrane cutoff for good retention), some carryover still occurred. Ideally this would not be an issue, as the peptide was present in excess during the reaction, but any excess unreacted oligonucleotide would be retained by the membrane. Purification by RP-HPLC used the solvent system of 0.1M TEAA and CH_3CN. When purifying, we looked for differences in retention time between the oligonucleotide and the conjugate. Purification was not trivial, and only a small amount of desired product was isolated by this method. Spin column purification was the method of choice for quick purification before analysis using *in vitro* splicing assays. Synthesized conjugates were diluted for analysis as described in Chapter 4.

2.2.2.2 Diels-Alder Conjugation

The Diels Alder cycloaddition is a widely utilized conjugation reaction that consists of two pieces – a conjugated diene and a dienophile, which undergo a rearrangement to create two new carbon-carbon bonds and a stable linkage. This reaction is advantageous in that it can proceed at room temperature or slightly elevated temperature under the right buffer conditions, forming a stable conjugate that is resistant to cleavage except under harsh conditions. Of the possible diene-dienophile combinations, we chose to use a furan ring as the diene and maleimide as the dienophile (**Scheme 2.3**). This combination has been utilized in the past for bioconjugation reactions with good result.[23] As there is not a diene-modifier for oligonucleotide synthesis, we purchased a furan-containing acid that could be easily coupled to an antisense oligonucleotide via amide bond formation. The maleimide had already been synthesized as described above. The reaction could proceed at room temperature in buffer conditions described previously, yielding the desired product.



Scheme 2.3

For our furan diene, we chose 3(2-furyl)propionic acid, a small molecule that would provide some flexibility between the peptide and oligonucleotide and provide the diene necessary for Diels-Alder coupling. The molecule was attached to an oligonucleotide containing a 3'-amino modifier via standard amide bond coupling using PyBOP as the coupling reagent. Attachment of the linker was moderately efficient, and an increase in temperature or reaction time at room temperature ensured that complete conversion would occur. If complete conversion did occur, the product was used without purification.

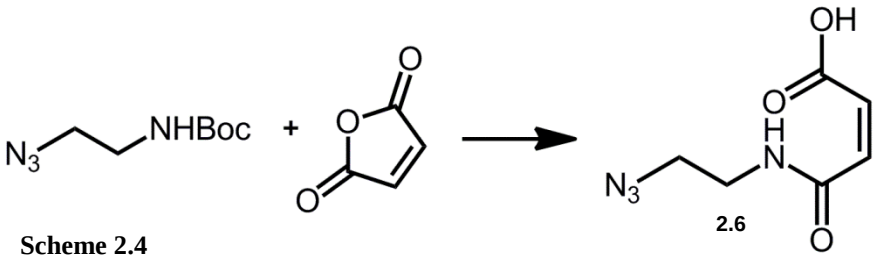
The Diels-Alder reaction had been shown previously to occur at room temperature under aqueous conditions, and we attempted this reaction under these conditions using peptides that had been previously synthesized for the maleimide-thiol addition.[24,25] Our results were not encouraging, as multiple reaction attempts did not produce a large quantity of product. Attempts to optimize the reaction focused on several areas. Molecule solubility was explored, as both oligonucleotide and peptide may have suffered from some insolubility under aqueous conditions due to the presence of side products or coupling reagents. Addition of CH_3CN to the reaction mixture assisted in increasing solubility. Reaction temperature was also explored, as it was postulated that the relatively large size of the two starting materials may have decreased coupling efficiency. Increasing reaction temperature to 37°C appeared to increase the amount of product produced, but temperatures higher than this were not explored due to the possibility of breakdown of the Diels-Alder adduct at high temperatures. Further improvements to the reaction may have involved using a size-exclusion agent such as polyvinyl alcohol (PVA) to increase effective concentration of the starting materials, or a chaotropic agent such as SDS or Triton-X that would have assisted in the breakdown of any aggregates of the two starting materials. A small number of conjugates were synthesized by this method, but the reaction was abandoned because of the higher coupling efficiency of the maleimide-thiol reaction.

2.2.2.3 Copper Catalyzed Azide-Alkyne Cycloaddition ('Click Chemistry')

The development of 'click chemistry' goes back to the term coined by Sharpless in 2001, defining a set of characteristics for orthogonal reactions that undergo complete conversion under mild conditions.[26] This definition was related to the simultaneous development by Sharpless and co-workers and Meldal and co-workers of the copper catalyzed azide-alkyne cycloaddition (CuAAC) that uses a copper catalyst to attach the two moieties together as a 1,2,3-triazole heterocycle. This reaction, first described by Huisgen, has become widely utilized as a means of conjugating different biomolecules together, and variations of the reaction using other substrates such as strained cyclooctynes have seen use as *in vivo* conjugation substrates by researchers such as Bertozzi and co-workers.[27]

Previous work done on this project using the click reaction for the synthesis of dendrimers (see Chapter 3) indicated that this reaction may be a viable alternative to the maleimide-thiol addition. The first issue was the development of a linker that could be used in

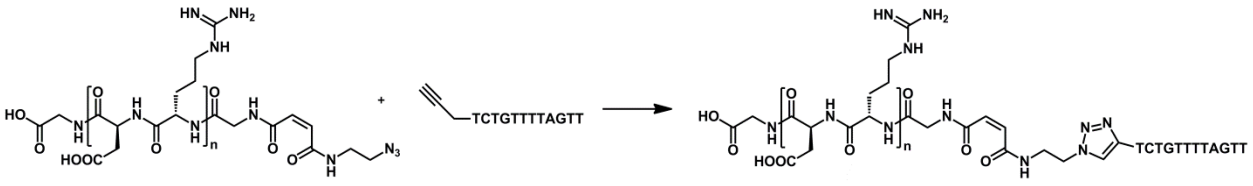
solid-phase peptide synthesis. Ideally, this linker would have a free carboxylic acid moiety for coupling to the N-terminus of the peptide. We had previously synthesized a Boc-protected 2-azidoethylamine for coupling to the C-terminus of our dendrons, but further manipulation of this molecule was necessary. With the synthesis of the maleimide linker in mind, we envisioned a similar synthesis using this azide linker rather than glycine as one of the substrates (**Scheme 2.4**). While not as efficient as the addition of glycine to maleic anhydride, this reaction proceeded fairly well, providing a reasonable amount of azide linker without the need for purification. This molecule could then be attached to the N-terminus of the peptide as the last step of SPPS. Cleavage from the resin and isolation of the product yielded the peptide containing the desired linker.



Scheme 2.4

Conjugation of the peptide and oligonucleotide was conducted under standard conditions for CuAAC (**Scheme 2.5**). The oligonucleotide was decorated with a commercially available 3'-alkyne moiety. Standard reaction conditions contained catalytic amounts of copper sulfate and sodium ascorbate, a reducing agent to assist in the reduction of the copper ion from Cu^{2+} to Cu^{+}, necessary for the reaction to occur. Different buffer conditions containing some combination of H_2O and a water-soluble organic solvent, such as DMF or *t*-BuOH, were used to increase the solubility of the starting materials and aid in the conjugation process. Reactions run at room temperature for 12 hours showed a modest conversion of reactants to products, and a number of changes to the reaction conditions were sought to improve yield. The relatively high boiling points of the solvents and stability of the starting materials to high temperatures suggested that an increased amount of heat would be useful in increasing reaction yield, and we utilized several methods to heat the reaction. First, the use of a microwave reactor was employed to increase the energy available for reaction to occur, as well as to decrease reaction times. Heating of the reactants in a microwave reactor or even a standard microwave for intermittent periods increased the yields substantially, with the only issue being loss of solvent in the standard microwave. A

second strategy utilized a thermal cycler, which has the benefit again of providing high temperatures for the coupling to occur. Heating to near boiling point in H_2O:DMF also increased yields, providing conjugates that could be analyzed in splicing assays.



Scheme 2.5

Purification of the CuAAC is not trivial. Initial attempts to purify the molecule by RP-HPLC were impeded by the fact that copper was not soluble in the HPLC solvents, leading to a buildup of excess copper on the HPLC guard column. This problem was addressed in a number of ways. Filtration of the solution through a 0.2 μm filter may have assisted in removing excess copper salts. A different filtration method uses the Amicon Ultra-0.5 mL Centrifugal Filter columns. An aqueous solution of the products and starting materials can be spun through the column, allowing the much smaller and water soluble copper sulfate and sodium ascorbate to pass through the membrane, leaving the desired product in the reservoir. This would not only help to remove excess copper, but would also help to purify the compound by allowing some of the excess starting materials to pass through the membrane. In doing so, HPLC purification was possible, yielding the desired conjugate for splicing assay analysis.

2.3 Conclusions and Future Directions

A series of synthetic RS domains have been designed and synthesized. These molecules were designed to examine a number of properties of the RS domain, including overall charge, charge distribution, and structural presentation of charged residues. Both α- and β-peptides have been synthesized and verified by ESI-MS and MALDI-MS, and a number of these molecules have been purified by RP-HPLC. Although purification was not easy, optimization of purification conditions yielded several peptides for conjugation. Several conjugation strategies were developed for connection of peptides and oligonucleotides, each having its own issues associated with synthesis and purification. Each conjugation scheme yielded several peptide-

oligonucleotide conjugates that were analyzed via *in vitro* splicing assays as discussed in Chapter 4.

There are a number ways in which the synthesis and conjugation of these activators could be improved. One issue is the synthesis of the peptide RS domains – as they are highly charged peptides, they are difficult to purify, and early attempts to purchase them from local sources were unsuccessful because they felt that the peptides could not be purified properly prior to conjugation. Subsequent searching has identified several companies, such as Sigma Aldrich and New England Peptide, which appear to have much more success at synthesizing the peptides. Rather than spending a great deal of time on synthesis, ordering the peptides at the onset of the project would have saved time. The bioconjugation reaction is another issue that should be addressed further. While the reactions chosen were fairly simple in theory, a great deal of work was needed to optimize them for these particular substrates. Use of a commercially available linker, while more expensive, may have solved this issue faster, as these reagents are usually optimized prior to being marketed as a coupling agent. Further optimization of purification methods for both the starting materials and conjugation products would likely yield better means of purifying these compounds, assisting in the ease of synthesis and isolation of more products.

Use of a different oligonucleotide could also be explored. Several other possible sites for hybridization were envisioned after reviewing the work of Hua et al., including targeting the Htra2β1 site and the 3' end of the exon (**Figure 2.5**).[7] These sites differed from the original binding site of the first SR protein mimics, and would have provided an opportunity to see what effect the RS domain would have in opposition to the negative effects of antisense oligonucleotide binding. Further studies on this project would benefit from analysis of these hybridization domains, and these molecules should be easily synthesized as the design of these molecules is modular.

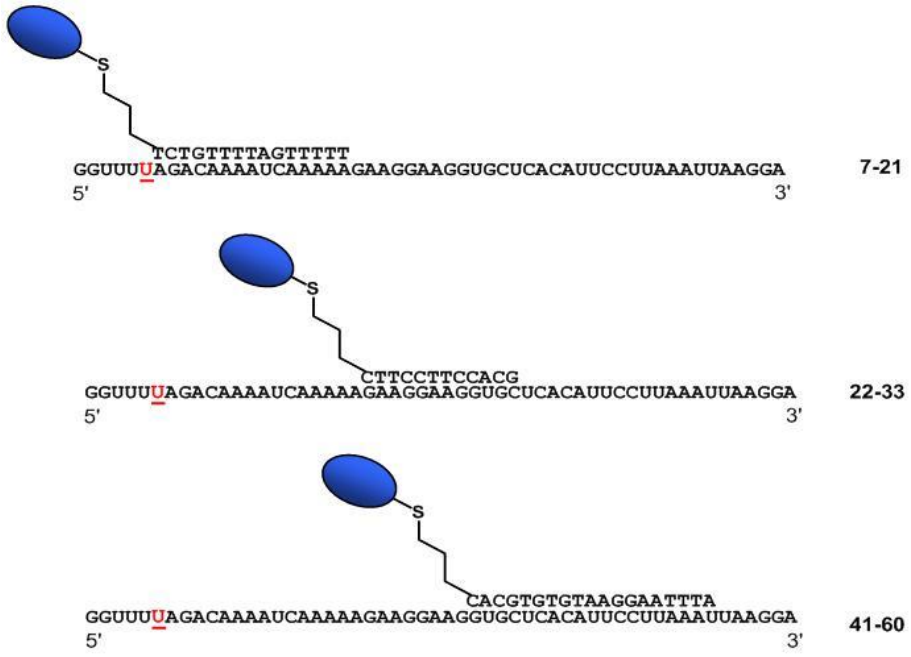


Figure 2.5: Potential Antisense Oligonucleotide Binding Sites

In addition to DNA oligonucleotides, a number of antisense molecules could be chosen in future work. The most commonly utilized molecules for antisense studies involve some sort of 2'-OH protected RNA oligonucleotides, including 2'-OMe and 2'-OMOE (methoxy ethyl), which impart stability to proteases and help to increase affinity for the sense strand of the mRNA. Other designs with varying properties could be explored, including morpholino oligonucleotides, locked nucleic acids, and peptide nucleic acids (PNA). Krainer and co-workers and Caputi and co-workers utilized PNA for their SR protein mimics, as it could be synthesized using the same reagents as the peptide on a single resin. Concerns with cost and availability of PNA deterred us from using it as the antisense nucleic acid in our initial designs, but further work on this project and exploration of alternatives to DNA would direct us to look again at whether or not PNA represents a possible alternative. Re-examination of the market for PNA resources may provide commercial sources for PNA that would be more reasonable to purchase. This would increase the ease of synthesis, as the peptide could be built directly from the PNA without the need for a bioconjugation method.

2.4 Materials and Methods

General

All reagents were purchased from commercial sources and used without further purification unless noted. Air and moisture sensitive reactions were performed in flame-dried glassware under an inert atmosphere of dry nitrogen. Dimethylformamide (DMF) was stored under nitrogen and over 4 Å molecular sieves. Tetrahydrofuran (THF) was distilled from sodium benzophenone ketyl, and *N*-methylmorpholine and CH_2Cl_2 were distilled from calcium hydride before use. All amino acids were dried by azeotropic distillation with dry DMF.

Thin layer chromatography was performed on pre-coated silica gel plates (silica gel 60, F_{254}) from EMD or Analtech, and eluting solvents are reported as volume ratios. Compounds were visualized using short wave UV light, iodine vapor, ceric ammonium molybdate (CAM), $KMnO_4$, or ninhydrin stains. Azide containing compounds were visualized with ninhydrin after treating with a 1% solution of triphenyl phosphine in methanol. Flash column chromatography was performed using Silica-P Flash Silica gel from Silicycle (40-63 μm particle size).

^1H NMR spectra were obtained in $CDCl_3$ or DMSO-d_6 using Varian Unity 400 MHz or 500 MHz spectrometers. Chemical shifts (δ) are reported in ppm using residual solvent protons as a reference (7.26 ppm for $CDCl_3$ and 2.50 ppm for DMSO-d_6). Mass spectra were obtained at the Mass Spectrometry Facility, University of Illinois. Low resolution ESI mass spectra were obtained using Waters Quattro II instrument. MALDI mass spectra were obtained using Applied Biosystems Voyager-DE STR and Bruker Daltonics UltrafleXtreme MALDI TOFTOF instruments. Solid-phase peptide synthesis was performed manually in a specially made reaction vessel. The loading of the resin used in all syntheses was 0.8 mmol amine/g resin. All peptide coupling reactions were monitored using the Kaiser test, low-res ESI-MS, and low-res MADLI-MS. Peptides were cleaved and deprotected using trifluoroacetic acid cleavage mix (87.5:5:5:2.5 TFA:DTT:H_2O:TIPS).

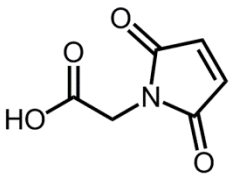
Peptide Synthesis General Procedure

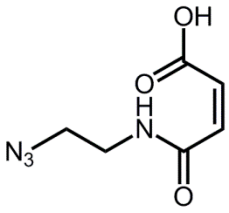
To a stirred solution of dried Fmoc-Asp(OtBu)-OH (50.0 mg, 0.12 mmol, 3 eq.) in dry DMF (1 mL) was added PyBOP (63.2 mg, 0.12 mmol, 3 eq.), HOBt-H_2O (16.4 mg, 0.12 mmol,

3 eq.), and DIPEA (28.2 μL, 0.16 mmol, 4 eq.). This solution was transferred to a solid phase peptide synthesis (SPPS) vessel contained pre-swelled 2-chlorotrityl resin loaded with glycine. The resin and peptide/coupling mixture were gently mixed with N_2 gas bubbling through the vessel. The reaction was monitored by ninhydrin test[28] and mass spectrometry**.** When the ninhydrin test indicated the reaction was complete, the vessel was drained by vacuum filtration, and beads rinsed with 2 mL DMF three times. Fmoc deprotection was accomplished by the addition of 2 mL 20% piperidine in DMF and mixing with N_2 gas three times, washing the beads with DMF between additions. The ninhydrin test was also used to determine when reaction had gone to completion. Subsequent amino acids were coupled in the same manner until the full peptide was synthesized.

Peptide Product Characterization

For each completed peptide, a small sample of resin was transferred to a glass vial with 1 mL 20% HFIP in CH_2Cl_2 and allowed to stand for 1 hr. Samples were then filtered through a glass-wool plugged pipette into a second glass vial, concentrated *in vacuo*, and submitted for low-res ESI-MS or low-res MALDI-MS. The following are completed peptides containing side-chain protecting groups, with MS technique indicated.

Linker: 

Azide-Linker: 

<u>Gly-(Asp-Arg)$_5$-NH$_2$</u> – LR-MS (ESI): calc. (M+H): 3195.9. Found (M+H): 3195.0.

<u>Gly-(Asp-Arg)$_5$-linker</u> – LR-MS (ESI): calc. (M+H): 3110.7. Found (M+H): 3110.2.

<u>Gly-(Asp-Arg)$_5$-Gly-linker</u> – LR-MS (MALDI): calc. (M+H): 3167.8. Found: (M+H): 3172.3.

<u>Gly-(Asp-Arg)$_5$-Gly-azide-linker</u> – LR-MS (MALDI): calc: (M+H): 3196.8. Found: (M+H): 3193.8.

<u>Gly-(Asp-Arg)$_6$-Gly-linker</u> LR-MS (MALDI): calc: (M+H): 3747.5. Found: (M+H): 3757.7.

<u>Gly-(Asp-Arg)$_8$-Gly-linker</u> – LR-MS (MALDI): calc. (M+H): 4906.9. Found (M+H): 4906.5

Gly-(Asp-Arg)$_{10}$-Gly-linker – LR-MS (MALDI): calc. (M+H): 6066.3. Found (M+H): 6066.7

Gly-(Asp-Arg)$_3$-Ala-(Asp-Arg)$_3$-Gly-linker – LR-MS (MALDI): calc. (M+H): 3818.5. Found (M+H): 3818.0.

Gly-(Asp-Arg)$_3$-Asp$_2$-Arg$_2$-(Asp-Arg)$_3$-Gly-linker – LR-MS (MALDI): calc. (M+H): 4906.9. Found (M+H): 4923.4.

Gly-(Asp-Arg)$_2$-Asp$_2$-Arg$_2$-Asp$_2$-Arg$_2$-(Asp-Arg)$_2$-Gly-linker – LR-MS (MALDI): calc. (M+H): 4906.9. Found (M+H): 4846.8.

Gly-(Asp-Ala)$_6$-Ala$_2$-Asp-Ala-Gly-linker – LR-MS (ESI): calc. (M+H): 2107.3. Found (M+H): 2107.2.

Gly-(Asp-Ala)$_8$-Gly-linker – LR-MS (MALDI): calc. (M+Na): 2229.0. Found (M+Na): 2229.0

Gly-(Ala-Arg)$_8$-Gly-linker – LR-MS (MALDI): calc. (M+H): 4105.9. Found (M+H): 4108.9.

Gly-(βAsp-βArg)$_4$-βAla-linker – LR-MS (MALDI): calc. (M+H): 2714.3. Found (M+H): 2715.9.

Gly-(βAsp-βArg)$_5$-NH$_2$ – LR-MS (ESI): calc. (M+H): 3336.1. Found (M+H): 3336.7.

Gly-(βAsp-βArg)$_5$-linker – LR-MS (ESI): calc. (M+H): 3251.0. Found (M+H): 3251.0.

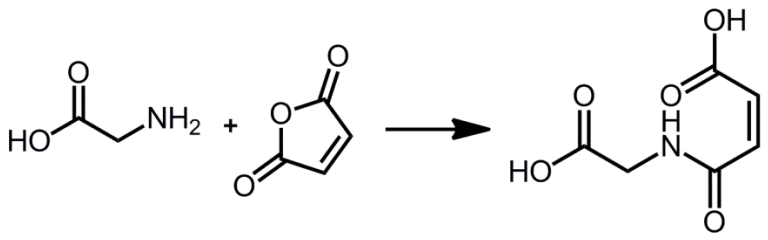
Gly-(βAsp-βArg)$_5$-βAla$_2$-linker – LR-MS (MALDI): calc. (M+H): 3393.1. Found (M+H): 3396.5.

Gly-(βAsp-βArg)$_6$-βAla-linker – LR-MS (MALDI): calc. (M+H): 4015.0. Found (M+H): 4015.3.

Gly-(βAsp-βArg)$_7$-βAla-linker – LR-MS (MALDI): calculated (M+H): 4537.6. Found (M+H): 4537.9

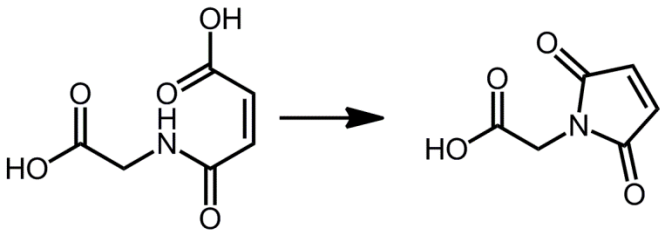
Gly-(βAsp-βArg)$_8$-βAla-linker – LR-MS (MALDI): calculated (M+H): 5145.3. Found (M+H): 5145.7.

Gly-(βAsp-βArg)$_{10}$-βAla-linker – LR-MS (MALDI): calculated (M+H): 6360.9. Found (M+H): 6368.7.



Glycine-maleic acid 2.1

To a stirred solution of maleic anhydride (981 mg, 10.0 mmol, 1 eq.) in acetic acid (5 mL) was added glycine (826 mg, 11.0 mmol, 1.1 eq.) in acetic acid (5 mL). After stirring for ~4 hours, a white precipitate was isolated by vacuum filtration. The solid was rinsed with three 5 mL portions of ice-cold H_2O, transferred to a round-bottomed flask, and dissolved in 90°C H_2O. Re-crystallization yielded 1.31 g of the maleic acid as a white solid that was isolated by vacuum filtration and dried under vacuum (7.57 mmol, 75.5% yield). ^{1}H NMR (400 MHz, DMSO) δ 9.19 (s, 1H, $CH_2NHC(O)$), 6.39 (d, 1H, *J* = 12.5 Hz, C(O)C*H*=CHC(O)), 6.28 (d, 1H, *J* = 12.5 Hz, C(O)C*H*=CHC(O)), 3.88 (d, 2H, *J* = 5.9 Hz, C(O)CH_2N)



Glycine-maleimide 2.2

Maleic acid (1.0 g, 5.8 mmol, 1 eq.) was dissolved in toluene (25 mL), and to this solution was added triethylamine (1.3 mL, 9.2 mmol, 1.6 eq.). Solution was heated to reflux under a Dean-Stark trap for one hour, and then cooled to room temperature. Removal of the solvent by rotary evaporation yielded orange oil, which was stored under vacuum overnight. 9 mL 1M HCl was added, and this solution was washed with three 15 mL portions of EtOAc. The organic layer was dried with Na_2SO_4, filtered, and solvent was removed by rotary evaporation to yield the maleimide-functionalized glycine as a yellow solid (749 mg, 4.8 mmol, 83.2% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.81 (s, 2H, CC*H*=C*H*C), 4.34 (s, 2H, C(O)CH_2N)

Maleimide-Thiol Conjugation 2.3

DTT Reduction: To a vial containing 400 µM SMN antisense oligonucleotide containing 3'-thiol modifier (100 µL) was added 1M DTT (20 µL). Solution was incubated at room temperature overnight. Aqueous layer was extracted three times with 200 µL EtOAc to remove excess DTT.

TCEP Reduction: To a vial containing 400 µM SMN antisense oligonucleotide containing 3'-thiol modifier (100 µL) was added 0.5M TCEP (20 µL). Solution was incubated at room temperature for 2 hours.

Conjugation: Aqueous layer containing oligonucleotide was transferred to a new vial containing 10 mM maleimide-containing peptide (200 µL) in 20% 0.1M triethylammonium acetate/CH_3CN (800 µL). Solution was incubated at 37°C overnight. Analysis by MALDI-MS indicated that the desired product was synthesized.

Maleimide-Thiol Product Characterization

All malemide-thiol peptide-oligonucleotide products were characterized by MALDI-MS. Products are listed as peptide-oligonucleotide type.

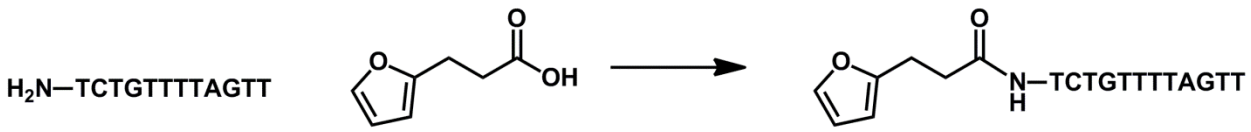
Gly-(Asp-Arg)$_5$-Gly-linker-Bcl-x – LR-MS: calc. (M+H) 5523.3. Found (M+H): 5603.8

Gly-(Asp-Arg)$_5$-Gly-linker-SMN – LR-MS: calc. (M+H): 5412.1. Found (M+H): 5412.0

Gly-(Asp-Arg)$_6$-Gly-linker-SMN – LR-MS: calc. (M+H): 5683.3. Found (M+H): 5687.2

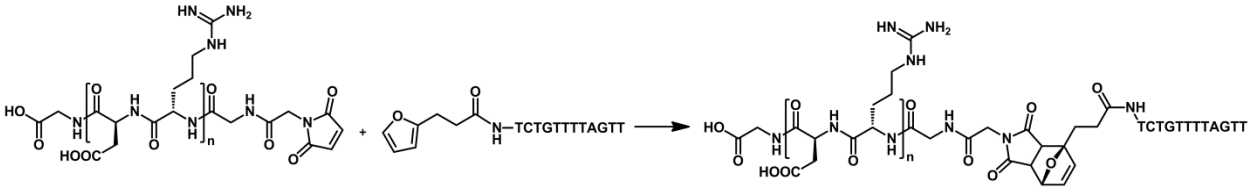
Gly-(Asp-Arg)$_3$-Ala-(Asp-Arg)$_3$-Gly-linker-Bcl-x – LR-MS: calc. (M+H) 5470.8. Found (M+H): 5379.5

Gly-(βAsp-βArg)$_5$-βAla$_2$-linker-SMN – LR-MS: calc. (M+H): 5396.2. Found (M+H): 5394.0



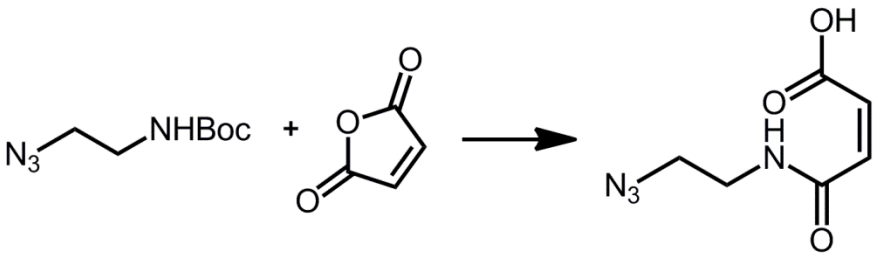
Furan Containing Oligo 2.4

To a stirred solution of 3(2-furyl)propionic acid (0.2 mg, 1.0 µmol, 5 eq.) in dry DMF (1 mL) was added PyBOP (0.7 mg, 1.0 µmol, 5 eq.) and DIPEA (0.4 µL, 3.0 µmol, 10 eq.). Solution was stirred at room temperature for 1 hr. Solution was transferred to a vial containing 400 µM SMN antisense oligonucleotide containing 3'-amino modifier (500 µL, 0.25 µmol, 1 eq.) and stirred at room temperature overnight. Analysis by MALDI-MS indicated that the desired product had been synthesized. MALDI-MS: calc. (M+H): 3981.5. Found (M+H): 3981.7.



Diels Alder Conjugation 2.5

To a vial containing $(DR)_5$ peptide in H_2O (200 µL, 10 µmol, 100 eq.) was added antisense oligonucleotide in DMF (200 µL, 0.1 µmol, 1 eq.). Solution was diluted to 500 µL with 0.1M phosphate buffer, pH 6.0 and incubated at 37°C overnight. MALDI-MS analysis indicated a small amount of the desired product. MALDI-MS: calc. (M+Na): 5630.1. Found (M+Na): 5639.1.



Azide Peptide-Oligo Conjugation Linker 2.6

To a stirred solution of Boc-protected 2-azidoethylamine (100.0 mg, 0.5 mmol 1.1 eq.) in CH_2Cl_2 (3 mL) was added trifluoroacetic acid (3 mL). Solution was stirred at room temperature under N_2 for 1 hr. Solvent was removed by azeotropic distillation and remaining solid was stored under vacuum overnight to remove excess TFA. Solid was resuspended in acetic acid (5 mL). To this solution was added maleic anhydride (47.9 mg, 0.5 mmol, 1 eq.) in acetic acid (5 mL) and solution was stirred at room temperature under N_2 for 3 hr. Following stirring, solvent was removed *in vacuo* and solid was resuspended in H_2O and acidified to pH 1 with 1M HCl. Aqueous layer was transferred to a separatory funnel and extracted three times with EtOAc (20 mL). Combined organic layers were dried over Na_2SO_4, filtered into a tared round bottom flask, and concentrated by rotary evaporation to yield the product as an oil (20.0 mg, 20.2% yield). ESI-MS: calc. (M-H): 183.2. Found (M-H): 183.1.

Azide linker was attached to $(Asp\text{-}Arg)_5$ peptide using solid phase synthesis as described above. Low-res MALDI-MS: calc. (M+H): 3194.47. Found (M+H): 3193.87.

Azide Peptide-Oligo Conjugation 2.7

To a stirred solution of peptide (400 µL, 0.4 µmol, 5 eq.) in 1:1 THF:H_2O (2 mL) was added SMN antisense oligonucleotide containing 3'-alkyne modifier (200 µL, 0.08 µmol, 1 eq.). To this solution was added 1M sodium ascorbate (4 µL, 4 µmol, 50 eq.) and 1M $CuSO_4$-$5H_2O$ (4 µL, 4 µmol, 50 eq.). Solution was heated in a microwave reactor at 50% power for ten 30 second intervals then stirred overnight. Analysis of the MALDI-MS indicated that the desired product had been synthesized. MALDI-MS: calc. (M+H): 5621.1. Found (M+H): 5572.4

2.5 References

(1) Kritzer, J. A.; Lear, J. D.; Hodsdon, M. E.; Schepartz, A. *J. Am. Chem. Soc.* **2004**, *126*, 9468.

(2) Kritzer, J. A.; Stephens, O. M.; Guarracino, D. A.; Reznik, S. K.; Schepartz, A. *Bioorg. Med. Chem.* **2005**, *13*, 11.

(3) Stephens, O. M.; Kim, S.; Welch, B. D.; Hodsdon, M. E.; Kay, M. S.; Schepartz, A. *J. Am. Chem. Soc.* **2005**, *127*, 13126.

(4) Walensky, L. D.; Kung, A. L.; Escher, I.; Malia, T. J.; Barbuto, S.; Wright, R. D.; Wagner, G.; Verdine, G. L.; Korsmeyer, S. J. *Science* **2004**, *305*, 1466.

(5) Cartegni, L.; Krainer, A. R. *Nat. Struct. Biol.* **2003**, *10*, 120.

(6) Hua, Y.; Sahashi, K.; Hung, G.; Rigo, F.; Passini, M. A.; Bennett, C. F.; Krainer, A. R. *Genes. Dev*. **2010**, *24*, 1634.

(7) Hua, Y.; Vickers, T. A.; Baker, B. F.; Bennett, C. F.; Krainer, A. R. *PLoS Biol.* **2007**, *5*, e73.

(8) Hua, Y.; Vickers, T. A.; Okunola, H. L.; Bennett, C. F.; Krainer, A. R. *Am. J. Hum. Genet.* **2008**, *82*, 834.

(9) Madocsai, C.; Lim, S. R.; Geib, T.; Lam, B. J.; Hertel, K. J. *Mol. Ther.* **2005**, *12*, 1013.

(10) Porensky, P. N.; Mitrpant, C.; McGovern, V. L.; Bevan, A. K.; Foust, K. D.; Kaspar, B. K.; Wilton, S. D.; Burghes, A. H. *Hum. Mol. Genet.* **2012**, *21*, 1625.

(11) Graveley, B. R. *RNA* **2000**, *6*, 1197.

(12) Nikolakaki, E.; Drosou, V.; Sanidas, I.; Peidis, P.; Papamarcaki, T.; Iakoucheva, L. M.; Giannakouros, T. *Biochim. Biophys. Acta* **2008**, *1780*, 214.

(13) Xiao, S. H.; Manley, J. L. *EMBO J.* **1998**, *17*, 6359.

(14) Stewart, K. M.; Horton, K. L.; Kelley, S. O. *Org. Biomol. Chem.* **2008**, *6*, 2242.

(15) Hamelberg, D.; Shen, T.; McCammon, J. A. *Proc. Nat. Acad. Sci. U. S. A.* **2007**, *104*, 14947.

(16) Pelsue, S.; Agris, P. F. *J. Protein Chem.* **1994**, *13*, 401.

(17) Perutz, M. F. *Pharm. Acta Helv.* **1995**, *69*, 213.

(18) Kritzer, J. A.; Tirado-Rives, J.; Hart, S. A.; Lear, J. D.; Jorgensen, W. L.; Schepartz, A. *J. Am. Chem. Soc.* **2005**, *127*, 167.

(19) Gelman, M. A.; Richter, S.; Cao, H.; Umezawa, N.; Gellman, S. H.; Rana, T. M. *Org. Lett.* **2003**, *5*, 3563.

(20) El-Faham, A.; Subiros Funosas, R.; Prohens, R.; Albericio, F. *Chemistry* **2009**, *15*, 9404.

(21) Ede, N. J.; Tregear, G. W.; Haralambidis, J. *Bioconjug. Chem.* **1994**, *5*, 373.

(22) Shafer, D. E.; Inman, J. K.; Lees, A. *Anal. Biochem.* **2000**, *282*, 161.

(23) Vieyres, A.; Lam, T.; Gillet, R.; Franc, G.; Castonguay, A.; Kakkar, A. *Chem. Comm.* **2010**, *46*, 1875.

(24) Marchan, V.; Grandas, A. *Curr. Protoc. Nucleic Acid Chem.* **2007**, *Chapter 4*, Unit 4 32.

(25) Marchan, V.; Ortega, S.; Pulido, D.; Pedroso, E.; Grandas, A. *Nucleic Acids Res.* **2006**, *34*, e24.

(26) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 2596.

(27) Sletten, E. M.; Bertozzi, C. R. *Acc. Chem. Res.* **2011**, *44*, 666.

(28) Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. *Anal. Biochem.* **1970**, *34*, 595.

CHAPTER 3

DESIGN AND SYNTHESIS OF DENDRIMER SR PROTEIN MIMICS

3.1 Introduction

The RS domain of SR proteins is intrinsically disordered, suggesting that a linear peptide mimic may be optimal for activation of alternative splicing.[1] The amino acid residues associated with these domains tend to lack secondary structure, instead maintaining a more flexible form. This disorder is thought to be important for protein function, but exploration of this property has been limited. Design of SR protein mimics containing a globular structure similar to proteins will help to assess whether a linear, disordered structure in a functional RS domain is an absolute requirement for activity. Our goal is to examine whether the globular arrangement of amino acids on the periphery of the dendrimer can yield an effective splicing activator.

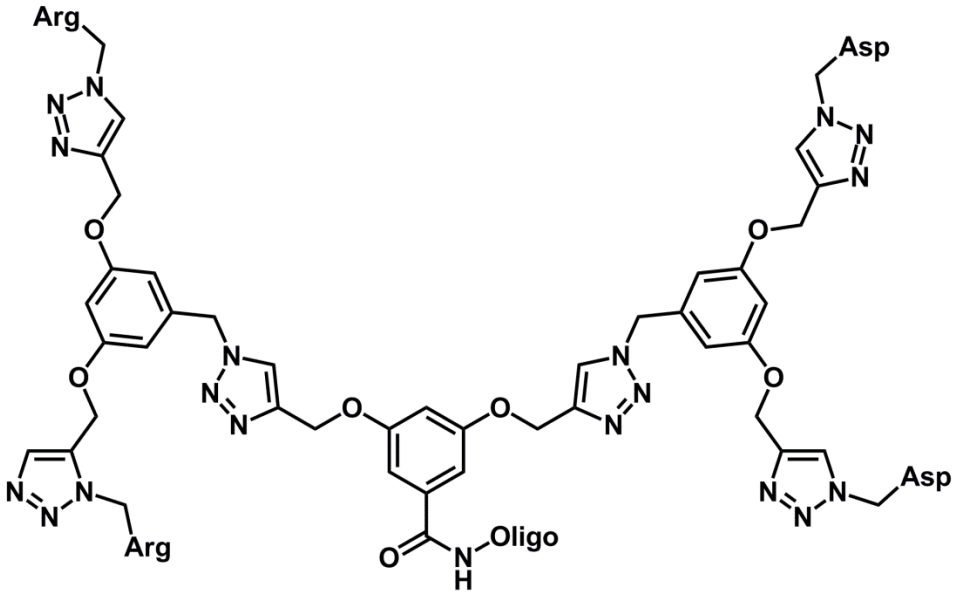
Dendrimers are branched polymers of well-defined molecular weight that contain three essential components: a central core, repeating branching units, and peripheral functionality. These components influence the dendrimer architecture, solubility, and molecular interactions, leading to a globular, multivalent structure.[2,3] Since their development in the 1980's, dendrimers have been applied to many biological applications, most notably drug and gene delivery.[4] A variety of dendrimer types have been developed that contain unique properties which are useful in a variety of contexts.

Peptide dendrimers, constructed with amide bonds, are well known for their structural similarity to globular proteins, reflected in their development as protein mimics.[5,6] The structural similarity of peptide dendrimers to proteins and protease resistance makes them ideal for use in a variety of biological applications and, like their linear counterparts, they can be easily synthesized using conventional SPPS methods. Although peptide dendrimers consist of a majority of our designs, several structural motifs were examined synthetically before our final design was analyzed by *in vitro* splicing, and a discussion of the evolution of our dendrimer motif is warranted.

3.2 Results and Discussion

3.2.1 First Generation Dendrimer – Click Dendrimer

The first-generation dendrimer RS domain mimic was designed based on the work of Sharpless and co-workers and utilized the Copper Catalyzed Alkyne-Azide Cycloaddition (CuAAC or 'click chemistry') to attach modified amino acids to a central aromatic core. In their work, Sharpless and co-workers developed a poly-aromatic dendrimer molecule that contained branching aromatic groups that were iteratively coupled via click chemistry.[7] These branching units could be modified to first be attached to the growing dendrimer then altered so that they can act as the branching unit to attach to a central core. In our design, we took advantage of the iterative attachment → modification → attachment scheme to develop a series of branching aromatic units that could be attached to a central core to give a globular structure.

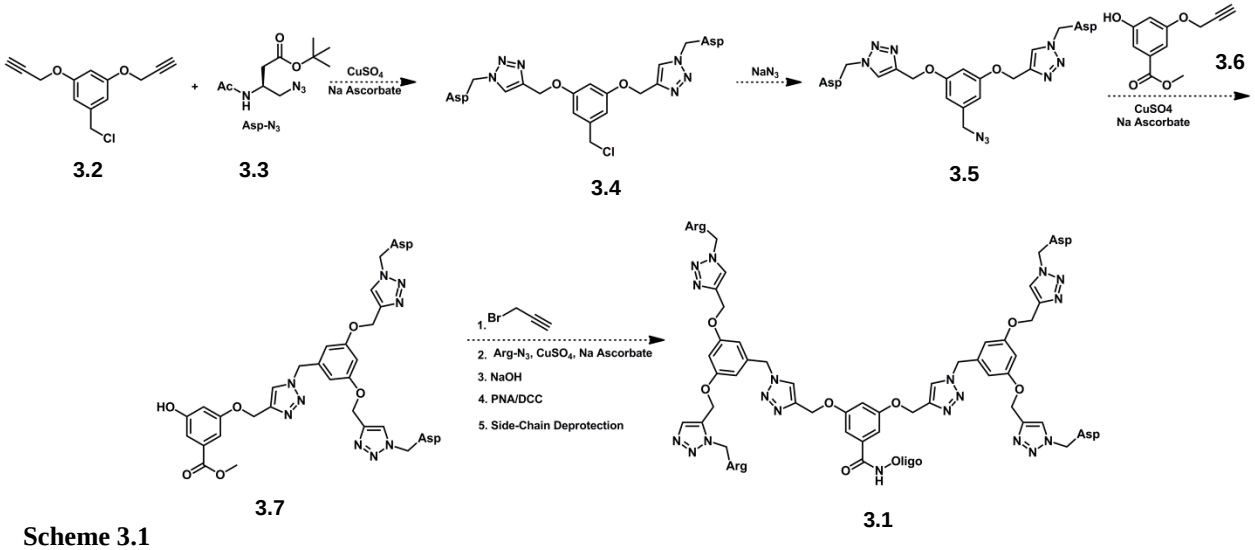


3.1

Figure 3.1. First Dendrimer Design

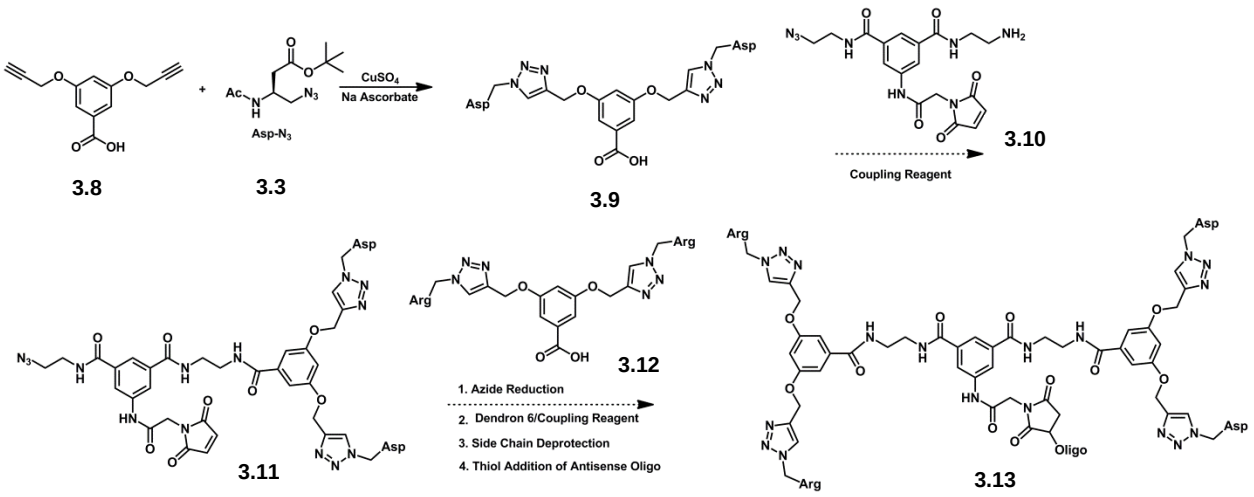
The desired goal for the dendrimer was to place the RS dipeptides on the periphery, so that they could recruit the splicing machinery and induce splicing. The initial design of the dendrimer structure is shown in Figure 3.1. In order for the amino acids to be available for coupling to the aromatic core, a series of manipulations were required. A priority during synthesis was keeping the amino acid side chain protected in order to prevent possible side reactions. We envisioned a reaction scheme in which the free acid moiety of the amino acid

would be modified through a series of reactions to produce an amino azide for the CuAAC to the central core. The aromatic core would be produced from a 3,5-dihydroxymethylbenzoate, as described by Sharpless and co-workers. Manipulation of the methyl ester of the aromatic core could lead to several possible outcomes. For example, reduction of the methyl ester would produce an alcohol moiety which could be used as a handle to produce an azide (**Scheme 3.1**). This handle could then be coupled to a second aromatic core via click chemistry, producing a full dendrimer.



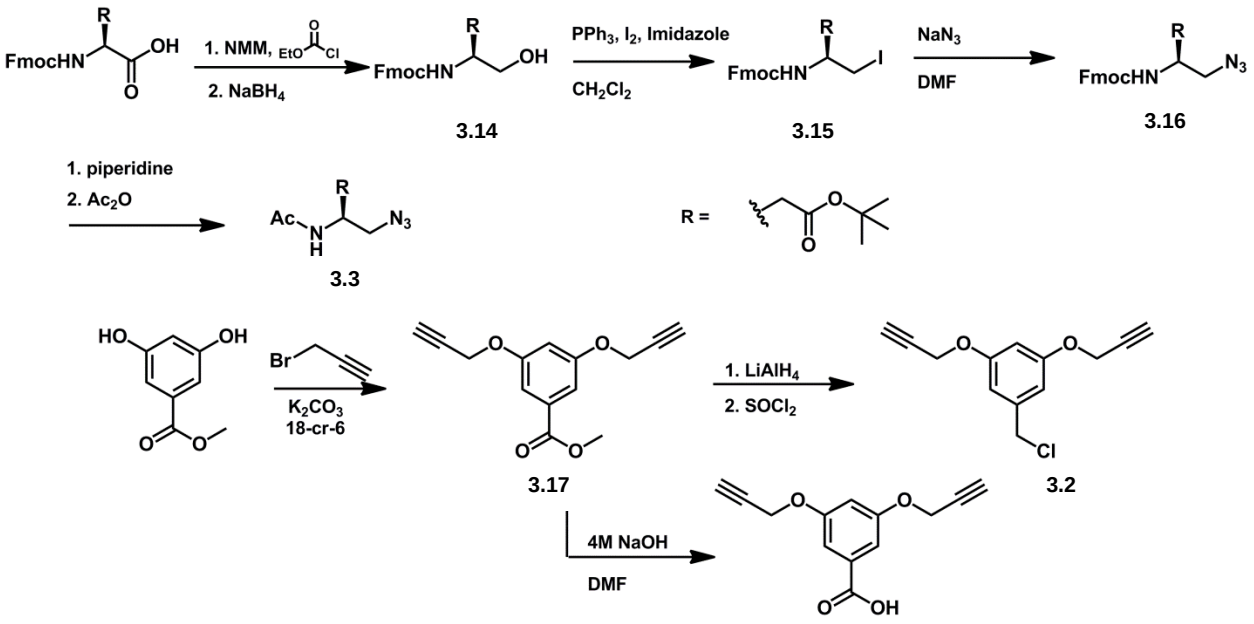
Scheme 3.1

A second possible method would be reduction of the ester to an acid, followed by attachment to a central core via amide bond formation (**Scheme 3.2**). Reduction of the other functionality followed by attachment of a second dendron would produce the dendrimer. Attachment of the dendrimer to the oligonucleotide would facilitated by a maleimide-thiol addition to the maleimide-containing branching unit attached to the core.



Scheme 3.2

Synthesis of the modified amino acids required several manipulations to add an azide moiety for attachment to the aromatic core (**Scheme 3.3**). First, the amino acid was converted to an activated mixed anhydride using ethyl chloroformate and N-methyl morpholine, which could then be reduced to amino alcohol **3.14** using sodium borohydride. This reaction was cumbersome, requiring extremely dry conditions and solvents to ensure that the anhydride was not hydrolyzed by excess water. Following isolation of the alcohol, several reactions were attempted to convert the molecule to an intermediate that could be used to make the azide. Use of thionyl chloride or directly converting the alcohol to an azide using diphenyl phosphorazidate and 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU) were unsuccessful.[8] Further analysis indicated that an iodide would be a suitable reactive intermediate for substitution with the azide. Reaction of the alcohol with triphenylphosphine, imidazole, and iodine resulted in substitution of iodide **3.15** in good yields. Following this reaction, simple stirring of the molecule with sodium azide in DMF yielded amino azide **3.16** in good yield. As a final step, the Fmoc protecting group of the amino azide was removed and substituted with an acetyl group (**3.3**), to prevent any non-specific interactions during subsequent couplings.



Scheme 3.3

Synthesis of the aromatic core was more straightforward (**Scheme 3.3**). Reaction of the 3,5-dihydroxymethylbenzoate with propargyl bromide and potassium carbonate produced the di-alkyne **3.17** in high yields. Several paths were envisioned with the methyl ester. First, it was reduced to an alcohol using lithium aluminum hydride, in order to be converted to the azide via a chloride intermediate. Synthesis of chloride **3.2** was not trivial, and only a small amount could be isolated. Subsequent attempts to synthesize the iodide instead of the chloride were also only moderately successful, and the azide was never isolated. A second pathway in which the methyl ester was reduced to acid **3.8** was more successful, and the acid was isolated in good yields.

The next step in the synthesis was attachment of the modified amino azides to the central core via CuAAC. A series of reaction conditions, varying the amounts of copper catalyst and solvents, were examined in order to determine the optimal reaction conditions for attachment. The products of the reaction were visualized by ESI-MS, and several produced good yields of the desired product that was purified by flash column chromatography. Although we were able to synthesize the aspartic acid-containing dendron, we abandoned this reaction scheme because it was too cumbersome for synthesis of reasonable quantities of the desired dendrimer.

There were a number of issues associated with the synthesis of this dendrimer. First, the conversion of the amino acids to amino azides was not trivial, requiring multiple reaction steps

that needed optimization and multiple attempts in order to be successful. In fact, the entire reaction scheme was only successful with aspartic acid. Repeated attempts with arginine were only moderately successful, and even then only the amino alcohol was isolated. Use of lysine instead of arginine was somewhat more successful, but even then only the amino iodide could be isolated. The attachment of the completed dendrons to the core was inherently challenging, as it required selective reduction of one side chain of the core while keeping the other in azide form. Different iterations of the core attempted to address this issue, but no design was settled upon that could simplify the attachment of two different components to the core. In addition, attachment of the maleimide to the core was difficult, and if it were completed prior to the final coupling step there was a chance that the ring would reopen in the presence of base. Finally, this design would only incorporate four charged residues on the periphery, and these would not even exist as dipeptides, making it extremely unlikely that we would see splicing activation. Thus, a different reaction scheme was sought.

3.2.2 Second Generation Dendrimer – Aromatic Polylysine Dendrimer

After spending a great deal of time optimizing the synthesis of the dendrimer, it became clear that the number of cumbersome manipulations and relatively limited potential for modular construction made this design too difficult to develop further. In particular, the difficulties with synthesizing and isolating both arginine and lysine amino azides indicated that this design would not be useful to pursue further. We hoped to find a design that required fewer cumbersome manipulations, a relatively straightforward synthetic pathway, and the possibility for substitution of different amino acids on the periphery.

After exploring a number of possible pathways, we settled upon a dendrimer design based on the work of Liskamp and co-workers.[9,10] This group has designed a number of different dendrimer motifs incorporating many types of functionality, including peptides and sugar moieties, as well as different types of connectivity between the branching units and central core. More importantly, these molecules have been used in a variety of biological contexts, from drug delivery to molecules that could be used to disrupt and block binding to cellular membranes.[11-14] These designs would present a viable starting point for our second generation dendrimers.

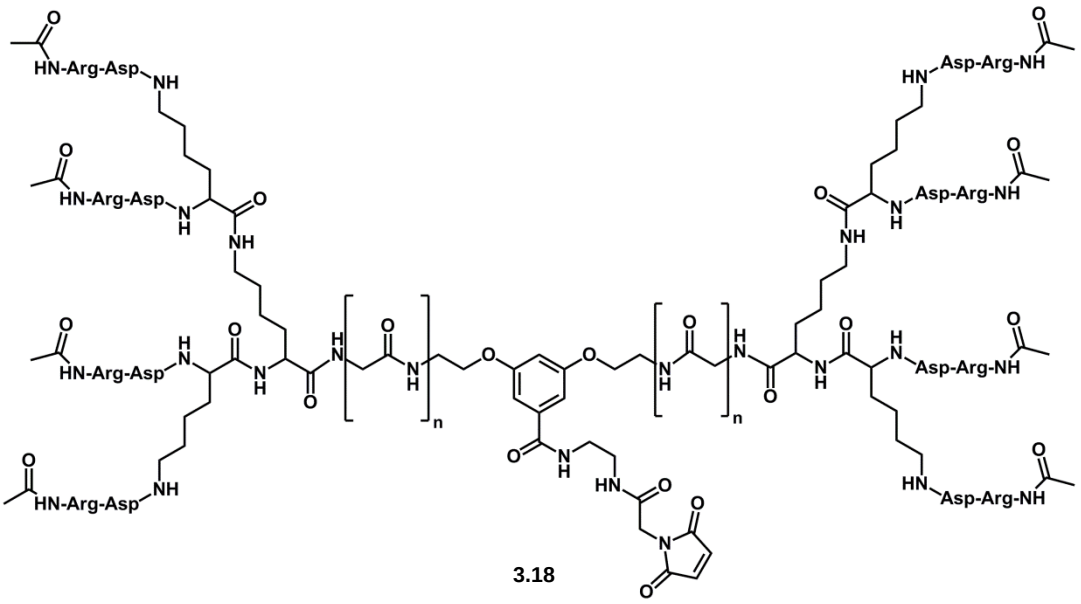
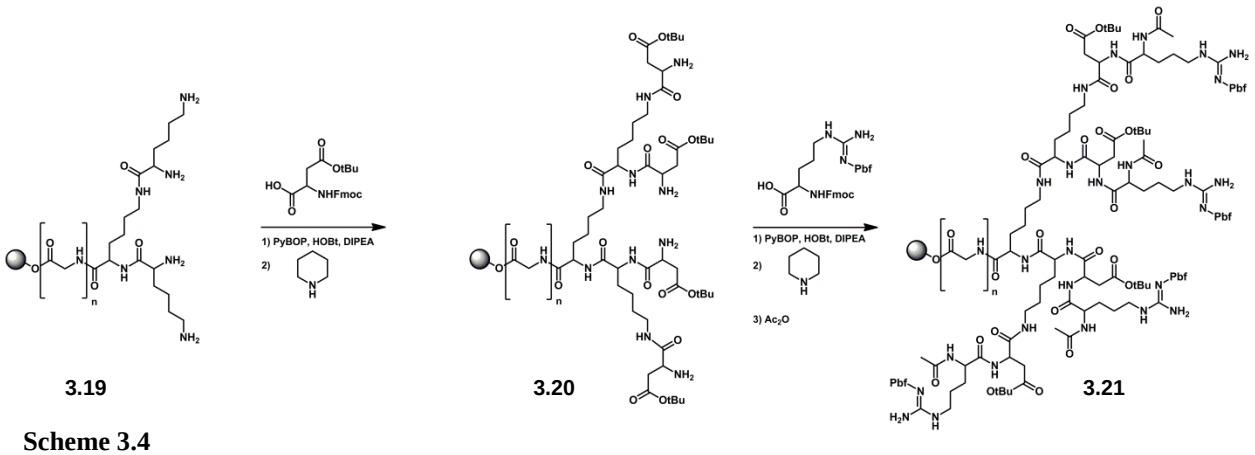


Figure 3.2. Second Dendrimer Design

We designed a series of dendrimers containing several components (**Figure 3.2**). The core of the dendrimer contains a trivalent aromatic motif with two identical attachment points for addition of the branching units, and a third, orthogonal attachment point for addition of the antisense oligonucleotide. This design uses the same aromatic core as the first generation dendrimer, but substitutes 2-bromoethylamine linkers for propargyl groups, to which can be attached the branching units via amide bond formation. The third site on the core is a modified version of the branch point from the first generation dendrimer, but instead of directly converting the starting methyl ester into an azide group, the ester can be reduced to an acid and then reacted with a linker molecule that attaches to the antisense oligo.

The dendrimer branches were redesigned to incorporate a polylysine branching dendron decorated on the periphery with the charged amino acids.[2] This design was advantageous over the previous design for several reasons. First, the dendron branches can be synthesized using solid phase peptide synthesis (SPPS) in a manner similar to the peptides, making them much simpler synthetically than the previously designed amino azides. Second, the number of residues on the periphery can be increased significantly compared to the previous dendrimer, from 2-4 amino acids to at least 8 dipeptide repeats on a full dendrimer. Third, the design is now modular, so that different combinations of amino acids can be added to the periphery from a single,

branching lysine dendron. Synthesis of the lysine dendrons is straightforward, using the same SPPS methods that were utilized for peptide synthesis (**Scheme 3.4**). Reaction times and equivalents of starting materials were increased as peripheral amino acids were added in order to ensure complete conversion. Terminal amino groups were acetylated upon completion of the synthesis to prevent undesired side reactions during attachment to the aromatic core. These dendrons were cleaved from the resin using HFIP to retain the side-chain protecting groups, again to prevent possible side reactions during attachment to the core.



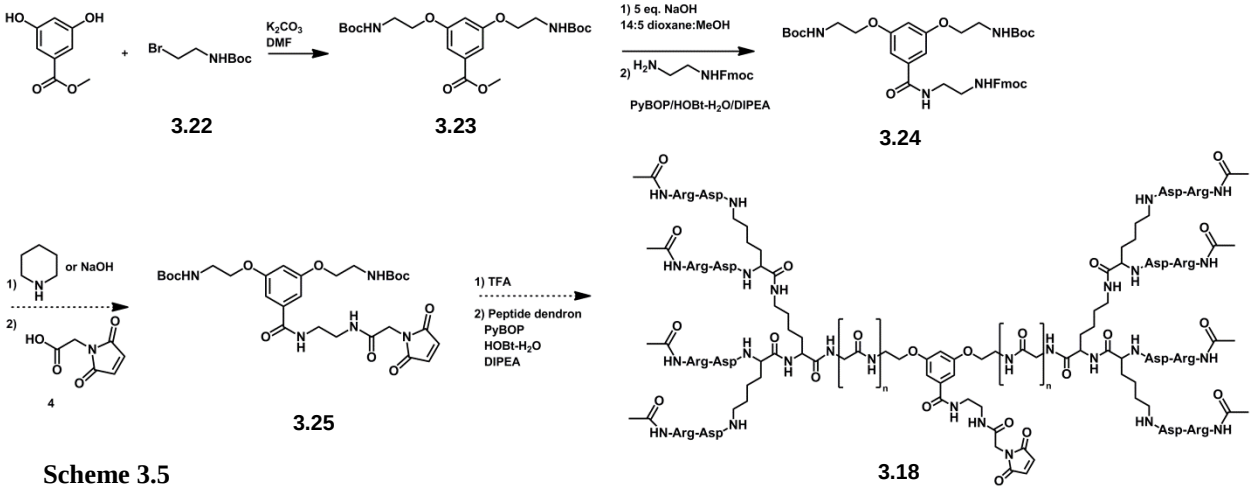
Scheme 3.4

In order to test the ability of the dendrimers to act as SR protein mimics, we synthesized series of peptide dendrons containing different numbers and arrangements of charged amino acids (**Table 3.1**). These include dipeptides containing aspartic acid, arginine, and alanine in different combinations. With this design, full dendrimers contain 8 dipeptide repeat motifs on the periphery. These dendrimers can be used to determine if molecules containing the DR dipeptide motif will be effective splicing activators, despite the non-linear presentation of amino acid residues. We have also designed and synthesized several peptide dendrons containing tetrapeptide motifs on the periphery, with the potential to contain 16 dipeptide repeats on the periphery of a single dendrimer. These molecules should be better splicing activators than the dendrimers containing dipeptide repeats, as the tetrapeptide motif has been shown to be effective at activating splicing in protein models.[15] Synthesis of these

Table 3.1. Synthesized Peptide Dendrons

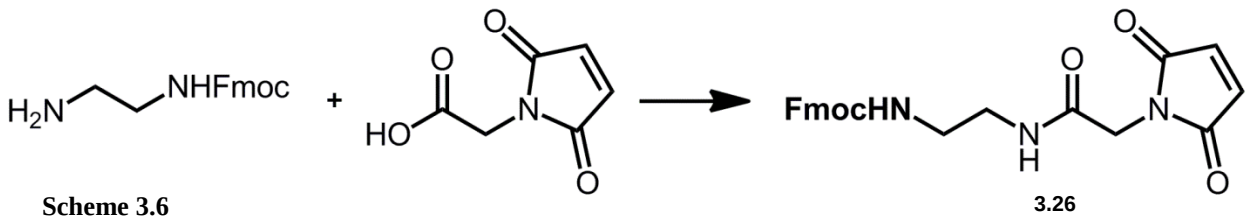
Peptide Dendrons
DR
DA
AR
AA
DRDR
DDRR

molecules required a large excess of reagents, as it was difficult to access multiple coupling sites with such large molecules.



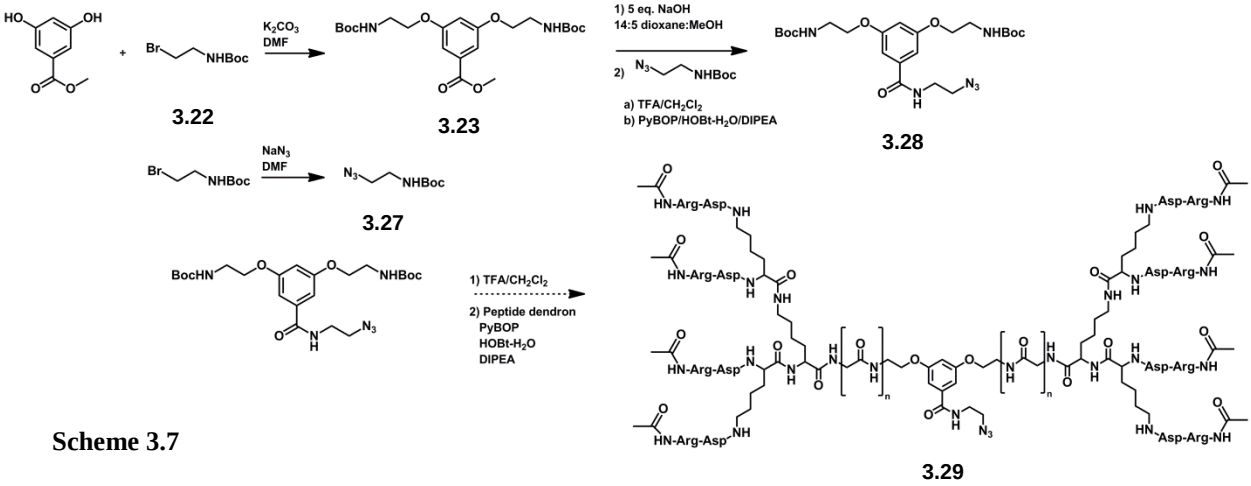
Scheme 3.5

The design and synthesis of the aromatic core went through several iterations as we determined the most effective dendrimer-oligonucleotide conjugation scheme. Our first design began the 3,5-dihydroxymethylbenzoate (**Scheme 3.5**). To synthesize the branching connection points, 2-bromoethylamine was first protected using di-tert-butyl dicarbonate and then attached to the free hydroxyl groups of the core to produce **3.23**. These reactions proceeded in good yields, with some impurities due to the unreacted di-tert-butyl dicarbonate. Subsequent saponification of the methyl ester and coupling of an Fmoc-protected ethylene diamine to give the aromatic core proceeded in moderate yields. The next steps included removal of the Fmoc group using piperidine and attachment of the maleimide linker discussed in Chapter 2, but this reaction proved cumbersome and isolation of the product was not completed (although the product was identified by ESI-MS). We attempted to circumvent this step by first attaching the maleimide linker to the Fmoc-protected ethylene diamine, which produced molecule **3.26** in moderate yields (**Scheme 3.6**), but this design was inherently flawed as the maleimide is sensitive to basic conditions. The Fmoc group must be removed by piperidine, and attempts to do so and couple with the aromatic core *in situ* were unsuccessful. An additional redesign was necessary.



Scheme 3.6 **3.26**

With our experience using CuAAC, as well as the many examples utilizing CuAAC as a bioconjugtion method, we explored using it as a method for attaching the dendrimer core to the oligonucleotide. To develop a handle for attachment, we substituted azide linker **3.27** for the Fmoc-ethylene diamine (**Scheme 3.7**). This would prevent unwanted problems with the maleimide linker and provide a point for attachment of the oligonucleotide. The azide linker was synthesized from the Boc-protected 2-bromoethylamine by nucleophilic substitution using sodium azide in DMF, in a similar fashion to our previous amino azide synthesis. This synthesis was straightforward and provided good yields of the product without the need for purification, as aqueous workup removed the excess sodium azide and sodium bromide salts. Removal of the Boc-protecting group using TFA and subsequent attachment to the core via amide bond formation provided aromatic core **3.28** in good yield.



Scheme 3.7 **3.29**

Attachment of the polylysine dendrons to the aromatic core was not trivial, and several attempts at redesigning the molecules to improve efficiency of coupling were moderate successful. To attach the dendrons to the core, the core was first deprotected with TFA to remove the Boc-protecting groups. Pre-activation of the dendrons using PyBOP and HOBt in the presence of diisopropylethylamine (DIPEA) was thought to be important to ensure that the coupling would occur in high yields. Upon mixing, the formation of a single-armed dendrimer

seemed to be straightforward and proceeded in good yields. Unfortunately, little to none of the desired full-sized dendrimer was isolated in our initial reaction attempts.

We speculated that there were a number of possible reasons for this lack of attachment. Running the reaction at room temperature was likely to not be energetically favorable and did not provide enough energy for the second attachment to occur. However, subsequent attempts using higher temperatures did not produce much more of the desired product. A more important issue was probably the large amount of steric bulk on the dendrons. The side-chains of the amino acids on the periphery remained protected to prevent unwanted side reactions, increasing the steric bulk substantially and likely preventing the second dendron piece from being able to access the core after the first had attached. We attempted to address this issue using several approaches (**Figure 3.3**).

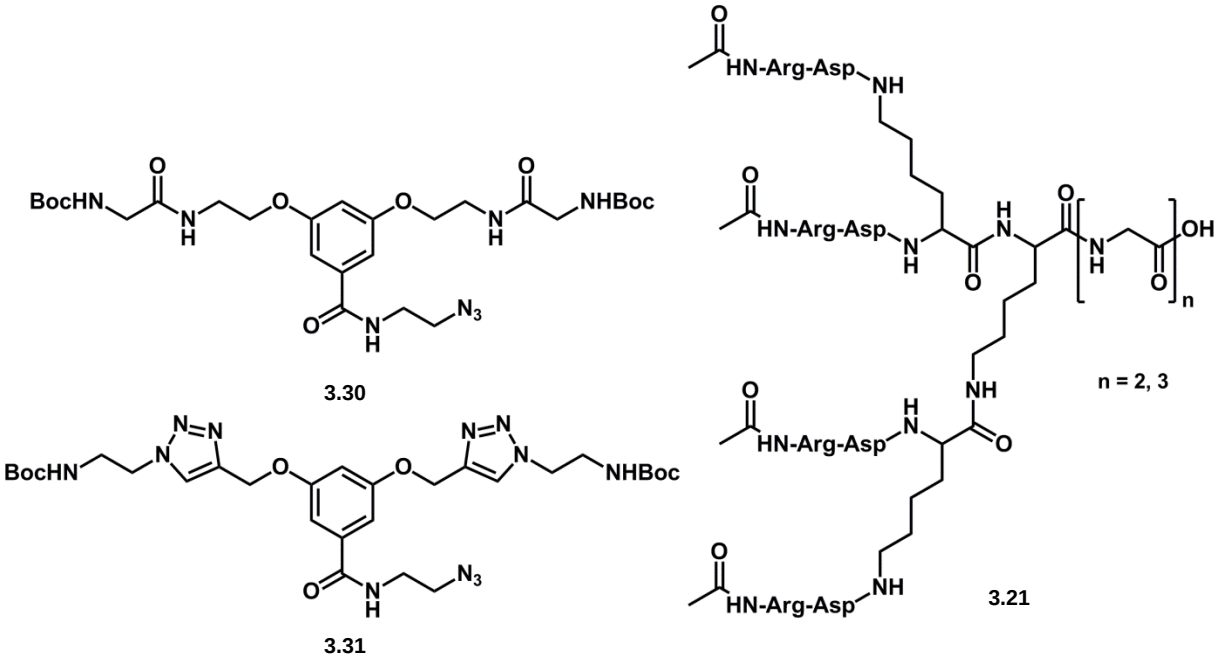


Figure 3.3. Extended Dendrimer Monomers

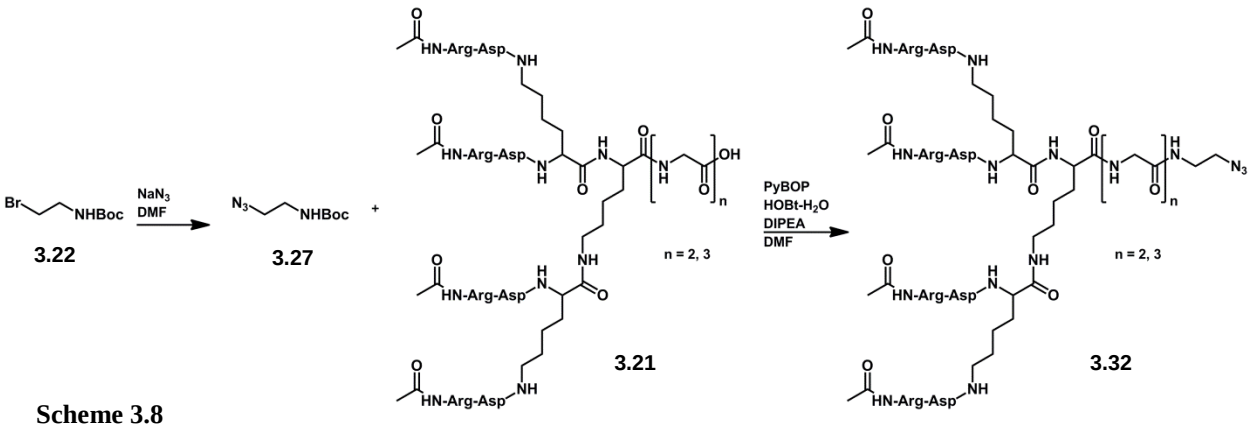
The first was to increase the length between the core and the dendrons while keeping the same attachment scheme for coupling the pieces together. Several methods were utilized to extend the branching units from the core, including attaching Boc-protected glycine (**3.30**) and using CuAAC to attach the Boc-protected 2-azidoethylamine (**3.31**). In addition, we began synthesis of dendrons containing 2-3 glycine residues instead of starting directly with lysine, again increasing the distance between the branching moieties and the core attachment. In both

cases, increasing the length did not seem to produce a more favorable result. A third possible issue was that the dendrons may have been undergoing aggregation due to the hydrophobic nature of the branching lysine residues. Use of a chaotropic agent such as SDS was employed to address this problem, but again there was no improvement in the yield of the desired product.

3.2.3 Third Generation Dendrimer – Polylysine Dendrimer

One of the difficulties associated with the synthesis of the full dendrimer as described in design 2 is that the peptide dendrons contain a great deal of steric bulk. As amide coupling was chosen to attach the dendrons to the central core, it was necessary to keep the side chains of the peripheral residues protected, increasing the steric bulk significantly. As a result, only a small amount of the full dendrimer could be synthesized, with a majority of product being a single-armed dendrimer. We addressed the issue producing a set of different molecules that increased the length between the central core and the branches, both on the branching dendron and on the core itself. Unfortunately, attempts to couple these pieces proved no more successful than with the original core. Redesign to incorporate the click reaction into the attachment rather than amide coupling (as utilized by Liskamp and co-workers in a number of papers) also did not prove successful (see Materials and Methods). Attempts to increase the yield of the full-sized product by using chaotropic agents such as sodium dodecyl sulfate (SDS) were also unsuccessful. Therefore, we sought a new strategy that still took advantage of the branching lysine dendrons without the issues associated with attaching the dendrons to the central core.

Examining the dendrimer design, we determined that the main issue appeared to be coupling of the dendrons to the core, and that there was potential for redesign at the point of attachment between the core and the branches. By simply repurposing the 2-azidoethylamine linker an attachment point on the branching dendron, the dendron itself could then be attached to the antisense oligonucleotide via CuAAC (**Scheme 3.8**). While this design decreased the number of charged residues, as well as the fully globular shape of the dendrimer in design 2, it would still be an interesting molecule series to examine, as the branching dendrons would provide a different arrangement of charged residues with which the spliceosome can interact.



Scheme 3.8

Synthesis of the branching dendrons was conducted as described above. Following synthesis, the dendrons were cleaved from the resin and isolated. We then coupled the dendrons with the 2-azidoethylamine that had been synthesized as the linker from the aromatic core to the oligonucleotide. This synthesis was straightforward, and produced a moderate yield of desired product **3.32** for coupling to the oligonucleotide. The only issues associated with this synthesis involved the aqueous workup of dendrons that contained alanine substitutions. Loss of product was occasionally an issue, and a modified aqueous workup that did not use strong acid or base ensured that the desired product was isolated. These dendron pieces were simple and straightforward to synthesize and isolate, making them much easier to work with than those molecules that had been designed previously.

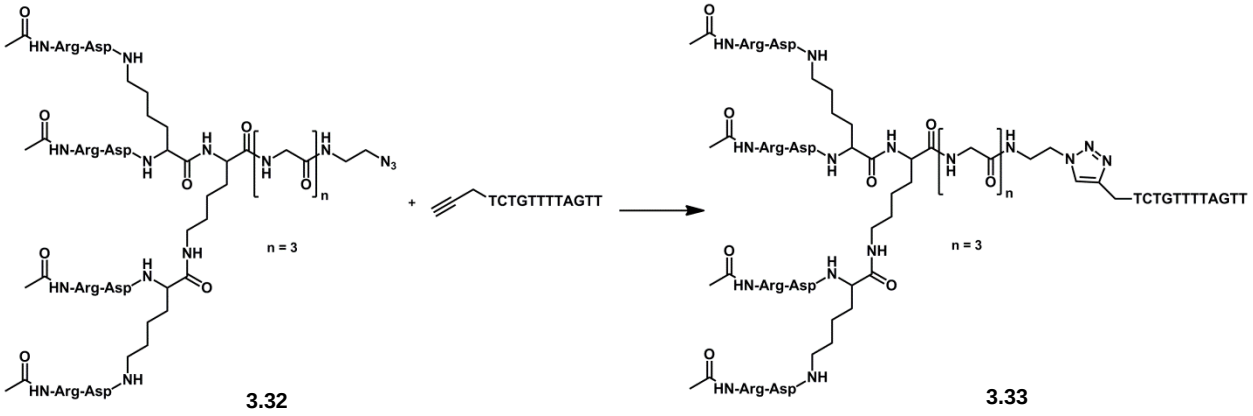
3.2.4 Dendrimer-Oligonucleotide Coupling

Coupling of the dendrimer to the antisense oligonucleotide was developed in parallel with the coupling of peptides and oligonucleotides discussed in Chapter 2, and a more thorough discussion of the coupling methods can be found there. Our initial dendrimers were designed to incorporate the maleimide ring onto the core for attachment of the oligonucleotide via maleimide-thiol addition. Unfortunately, these molecules were never synthesized, and no attempts were made to couple the molecules in this fashion.

3.2.4.1 Copper-Catalyzed Alkyne-Azide Cycloaddition

The next strategy using our second generation dendrimers was to utilize the CuAAC to attach the dendrimer core to the antisense oligonucleotide. Synthesis of the dendrimer core is described above, producing the azide-containing moiety for attachment. We were able to obtain

antisense oligonucleotides containing a commercially available 3'-alkyne modifier. As we were only able to obtain a single-armed dendrimer from our synthesis, the click reaction was attempted using that molecule. Unfortunately, no products were obtained from this reaction. Due to the relatively small amount of material, this dendrimer conjugation was not pursued further.



Scheme 3.9

Conjugation of the third generation dendrimer to the antisense oligonucleotide was much more successful than other iterations. As described previously, we utilized the same antisense oligonucleotide containing the 3'-alkyne moiety to conjugate to our azide-containing dendrons under the conditions for CuAAC described by Sharpless and co-workers and further developed by others (**Scheme 3.9**). The first attempts at this conjugation used catalytic copper sulfate and sodium ascorbate at room temperature, yielding a modest amount of the desired product and a good deal of starting material. Attempts to optimize the reaction examined reaction solvent, switching from H_2O:*t*-BuOH mixtures to H_2O:DMF mixtures to increase solubility of the starting materials. Different amounts of copper were also examined, which indicated that a larger than catalytic amount would be necessary to increase yields. Reaction temperature was varied as well, with use of microwave reactors, standard microwave ovens, and thermal cyclers all being used to increase the energy available for the reaction to occur. Optimization of these conditions yielded a number of molecules that could be used for splicing.

Purification of the dendrimers was not trivial. As described in Chapter 2, the biggest issue associated with purification was buildup of excess copper on the guard column when using RP-HPLC. Removal of this copper using filtration devices alleviated the load of copper on the

column, making it easier to purify the molecules by HPLC. Several other methods were attempted to purify these conjugates further, including membrane dialysis and filtration using the Amicon Ultra-0.5 mL Centrifugal filtration devices. Purification by dialysis led to precipitation of the reaction mixture into the dialysis membrane, and the products isolated by dialysis contained primarily the antisense oligonucleotide. Filtration through Amicon columns was much more successful, yielding a reasonable amount of the products for analysis. Further work using the CuAAC would benefit from additional purification methods. In particular, size exclusion chromatography would be the ideal choice for separating out the products from starting materials. Removal of the copper before purification would ensure that issues with column-clogging would be eliminated.

3.2.4.2 Alkyne-Cyclooctyne Conjugation

A recently emerging area in the field of bioconjugation is the use of strained cyclooctynes as the alkyne moieties for coupling to azides. Developed in great part by Bertozzi and co-workers, these molecules have been used in a number of systems both *in vitro* and, more importantly, *in vivo*, to synthesize biologically relevant molecules.[16-18] Some examples include the conjugation of fluorescent markers to cellular surfaces for visualization of cells, as well as ^{18}F-labeling of cancer imaging agents for analysis and testing of chemotherapeutics.[19] This reaction has been dubbed "strain-promoted click chemistry" because the strain of the cyclooctyne ring provides the energy necessary for the reaction to occur, and also "copper-free click chemistry" because, in the presence of this strain, a copper catalyst is not necessary for the reaction to occur. Our issues associated with the synthesis and purification of the CuAAC led us to investigate whether dendrimer-oligonucleotide conjugation using this method would be a viable option for synthesizing our molecules.

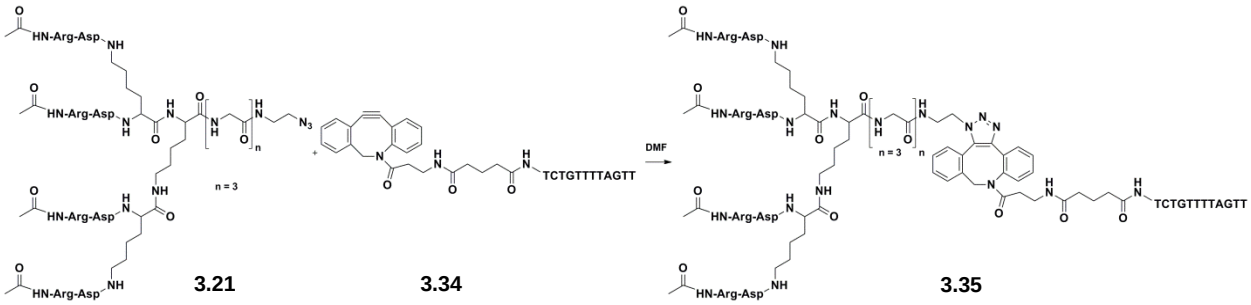
The field of strain-promoted click chemistry is relatively new, meaning that only a small number of commercial sources were available for the synthesis of our conjugates. Although recent advances have led to the development of nucleotide-based cyclooctynes that can be incorporated into DNA synthesis procedures such as PCR, few are available that can be used in conventional DNA synthesis machines.[20] More importantly, there are few commercially available 3'-modifiers containing the desired cyclooctyne moiety. A search of possible sources led us to a company called Click Chemistry Tools, which had a number of possible options. As

stated, no 3'-modifiers were available, although there exist cyclooctyne phosphoramidites that can be incorporated during 3'-5' DNA synthesis at the 5'-end. Unfortunately, our chosen oligonucleotide source (Yale) could not include these molecules in their synthesis, so a phosphoramidite was not the proper choice for our molecules. Instead, we were able to obtain a strained cyclooctyne linker with an activated NHS-ester for coupling to a free amino group. The 3'-amino modifier is readily available, and we obtained both the modified oligonucleotide and the cyclooctyne linker for coupling.

3.34

Scheme 3.10

Attachment of the activated NHS-ester cyclooctyne to the antisense oligonucleotide required optimization to produce desired product **3.34** (**Scheme 3.10**). Initial attempts to couple the starting materials used 1X phosphate buffered saline, pH 7.4, which was used in the protocol accompanying the molecule to couple the NHS-ester to a protein target. This reaction produced only a small amount of the desired product after reacting overnight, suggesting that the conditions needed to be improved. The pH of the reaction may not have been sufficiently high to make the amine nucleophilic, so the buffer was changed to 0.1M phosphate buffer, pH 8.0. Continued stirring at room temperature only increased the yield slightly. Taking our previous experience with coupling reactions into account, it was decided that the temperature of the reaction was to blame for our lack of success. By increasing the reaction temperature to 37°C using an incubator, and in later reaction to 94°C using a thermal cycler, we were able to get complete conversion of the antisense oligonucleotide to contain the strained cyclooctyne linker. One difficulty with using phosphate buffer for our reaction, however, was that the phosphate made analysis by MALDI-MS much more difficult. Removal of phosphate buffer salts by standard desalting methods such as Zip Tip is an alternative possibility for analysis, if phosphate must be used. Knowing that a base was necessary to drive the reaction, a different strategy was used in which a slight excess of diisopropylethylamine (DIPEA) was added to the reaction mixture in DMF. Mixing of this reaction using the thermal cycler led to complete conversion to the product, with the added benefit that the reaction could be easily monitored.



Scheme 3.11

Synthesis of the dendrimer-oligonucleotide conjugate using the strained cyclooctyne was conducted under several sets of conditions (**Scheme 3.11**). The conditions suggested by the manufacturer were to use a buffer in which the azide-containing molecule is soluble and then add the strained cyclooctyne, so we first used an H_2O:CH_3CN buffer, as it was used for previous syntheses. These reactions were moderately successful, producing only a small amount of the desired product. With our experience using the thermal cycler to attach the strained cyclooctyne linker, we attempted the reaction under the same conditions using DMF as the buffer because of its increased boiling temperature. Gratifyingly, this strategy was successful, producing desired product **3.35** with complete conversion of the oligonucleotide starting material. These molecules could be purified as described previously, yielding the desired product for splicing.

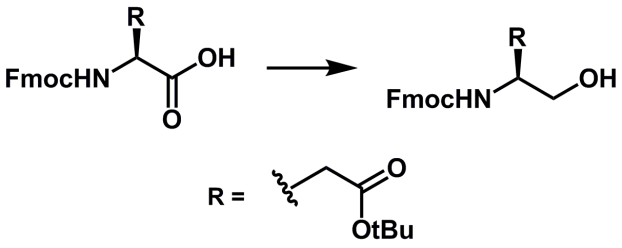
3.3 Conclusions and Future Work

The design and synthesis of dendrimer RS domains was accomplished after going through several iterations of design. An initial design using modified amino acids and CuAAC or 'click chemistry' produced several molecules, but proved too cumbersome a synthesis to pursue for production of the dendrimers. A second design based on the work of Liskamp and co-workers substituted the modified amino acids for polylysine dendrons, as well as modifying the central core for conjugation of the dendrimer to the antisense oligonucleotide via click chemistry. This design was much more successful and straightforward to synthesize, but suffered from issues with steric bulk associated with the final synthetic step. Repeated attempts yielded only a single-armed dendrimer. Further redesign that removed the aromatic core and directly attached the dendron to the antisense oligonucleotide via click chemistry proved to be the most straightforward, generating several molecules that could be purified and examined by *in vitro*

splicing assay. Purification of these molecules was not trivial, requiring several alternative methods to remove the copper from the reaction mixture prior to other purification strategies. Finally, an alternative method for conjugation utilizing strained cyclooctyne rings provided a second method for attachment of the dendrimer and antisense oligonucleotide that could be more easily purified and used in splicing assays, as discussed in Chapter 4.

Future synthesis and analysis of the dendrimer-oligonucleotide conjugates could benefit from the analysis of alternative conjugation methods. For example, different synthetic procedures for polypeptides attached to a central aromatic core via CuAAC or amide bond formation may provide additional methods for synthesizing the full dendrimer. Optimization of reaction temperature, solvent, and use of chaotropic agents could also improve the dendrimer synthesis. If the branching lysine dendron is utilized as the sole dendrimer piece, use of additional conjugation methods such as commercially available linkers may be useful in ensuring that the coupling occurs with high efficiency and that the final product can be easily purified. Alternative methods for purification of these molecules would also be beneficial, so that issues associated with methods such as RP-HPLC can be avoided. With some optimization, these molecules could be more easily synthesized, purified, and analyzed for splicing modulation activity.

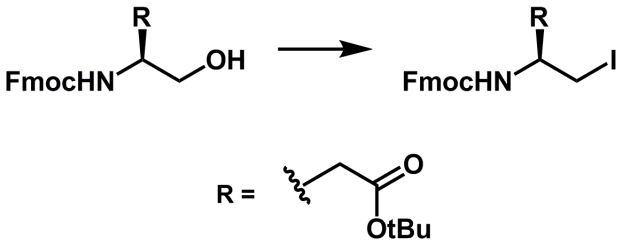
3.4 Materials and Methods



Amino Alcohol 3.14

To a cooled (0°C) solution of dry Fmoc-protected aspartic acid (2.0 g, 4.9 mmol, 1 eq.) in dry THF (5 mL) in a flame-dried Schlenk flask under argon was added *N*-methylmorpholine (641 µL, 5.8 mmol, 1.2 eq.) and ethylchloroformate (558 µL, 5.8 mmol, 1.2 eq.). The solution was stirred for 45 min at 0°C under argon and monitored by TLC (4:1 Hexane:EtOAc).

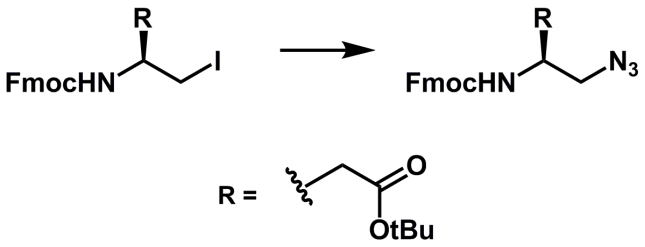
After stirring, the N-methylmorpholonium salts were filtered off through an oven-dried, cotton-plugged Pasteur pipette into a flame-dried Schlenk flask under argon using an oven-dried Pasteur pipette. The precipitate was washed with dry THF (10 mL) into the Schlenk flask. The combined filtrates were cooled to 0°C with an ice bath. To the solution was added a solution of $NaBH_4$ (276 mg, 7.3 mmol, 1.5 eq.) in H_2O (2.8 mL). The reaction was stirred at 0°C until H_2 evolution ceased (45 min), when TLC indicated that the reaction was complete. Excess $NaBH_4$ was quenched by the addition of 10 mL H_2O. The organic layer was washed with 20 mL sat. $NaHCO_3$. The aqueous layer was washed with three 20 mL portions of EtOAc, and the combined organic layers were dried over Na_2SO_4. The solution was filtered and concentrated *in vacuo*, followed by flash column chromatography (1:1 EtOAc:Hexane), to yield 1.41 g (3.6 mmol**,** 73.2% yield) of pure solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.76 (d, 2H, *J* = 7 Hz, Fmoc ar), 7.59 (d, 2H, *J* = 7 Hz, Fmoc ar), 7.40 (t, 2H, *J* = 7 Hz, Fmoc ar), 7.31 (t, 2H, *J* = 7 Hz, Fmoc ar), 5.50 (d, 1H, *J* = 8 Hz, N*H*), 4.40 (d, 2H, *J* = 6.8 Hz, Fmoc CH_2), 4.21 (t, 1H, *J* = 6.8 Hz, Fmoc C*H*), 4.11 (m, 1H, C*H*CH_2OH), 3.73 (t, 2H, *J* = 5 Hz, CH_2OH), 2.57 (m, 2H, *J* = 6 Hz, CH_2(O)O*t*Bu), 2.46 (t, 1H, *J* = 5.5 Hz, CH_2O*H*), 1.45 (s, 9H, -O*t*Bu)



Amino Iodide 3.15

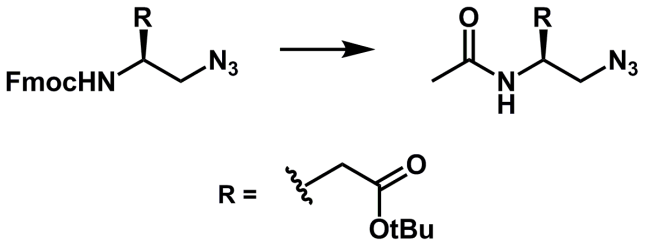
To a stirred solution of triphenylphosphine (396 mg, 1.5 mmol, 3 eq.) in dry CH_2Cl_2 (4 mL) was added solid I_2 (639 mg, 2.5 mmol, 5 eq.). After allowing I_2 to dissolve, imidazole (171 mg, 2.5 mmol, 5 eq.) was added, followed by the amino alcohol (200.0 mg, 0.5 mmol, 1 eq.). The solution was stirred for 3 h, when starting material was consumed as confirmed by TLC (5:1 EtOAc:Hexane). Solvent was removed *in vacuo* using a secondary trap, and flash column chromatography (5:1 EtOAc:Hexane) yielded 186.3 mg (0.37 mmol, 73% yield) of the iodide as a pure white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.78 (d, 2H, *J* = 7.6 Hz, Fmoc ar), 7.60 (dd, 2H, *J* = 7,1 Hz, Fmoc ar), 7.40 (t, 2H, *J* = 7,1 Hz, Fmoc ar), 7.32 (t, 2H, *J* = 7 Hz, Fmoc ar), 5.36

(d, 1H, *J* = 9 Hz, N*H*), 4.39 (2H, Fmoc CH_2), 4.23 (t, 1H, *J* = 7 Hz, Fmoc C*H*), 3.96 (m, 1H, C*H*CH_2I), 3.42 (m, 2H, CH_2I), 2.61 (ddd, 2H, *J* = 16,6 Hz, CH_2(O)O*t*Bu), 1.54 (s, 9H, -O*t*Bu)



Amino Azide 3.16

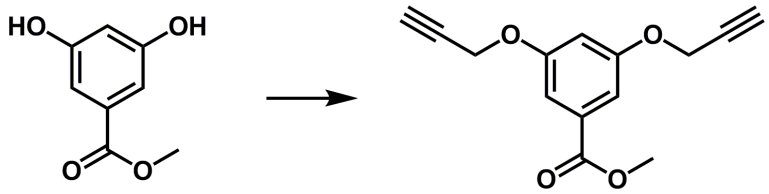
To a stirred solution of iodide (500 mg, 0.99 mmol, 1 eq.) in dry DMF (5 mL) was added NaN_3 (320 mg, 4.9 mmol, 5 eq.). The solution was stirred at room temperature for 5 h, when starting material was consumed as confirmed by TLC (2:1 Hexane:EtOAc). Solvent was removed *in vacuo* using a secondary trap, and flash column chromatography (2:1 Hexane:EtOAc) yielded 358 mg (0.85 mmol**,** 86.0% yield) of the azide as a pure white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.77 (d, 2H, *J* = 7 Hz, Fmoc ar), 7.59 (d, 2H, *J* = 7,1 Hz, Fmoc ar), 7.41 (t, 2H, *J* = 7 Hz, Fmoc ar), 7.31 (t, 2H, *J* = 7,1 Hz, Fmoc ar), 5.43 (d, 1H, *J* = 9 Hz, N*H*), 4.41 (2H, *J* = 17,10,7 Hz, Fmoc CH_2), 4.22 (t, 1H, *J* = 7 Hz, Fmoc C*H*), 4.14 (m, 1H, C*H*CH_2N_3), 3.51 (2H, *J* = 20,12,5 Hz, CH_2N_3), 2.53 (d, 2H, *J* = 6 Hz, CH_2(O)O*t*Bu), 1.46 (s, 9H, -O*t*Bu)



Acyl-protected Amino Azide 3.3

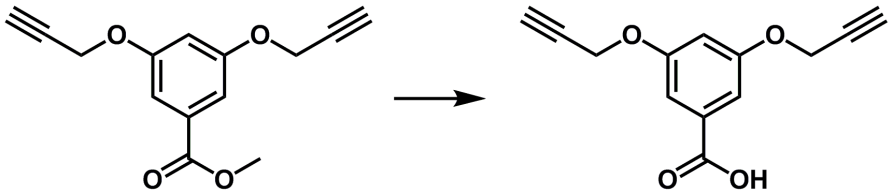
To a stirred solution of the Fmoc-protected azide (358 mg, 0.85 mmol, 1 eq.) in dry DMF (3 mL) was added 1 mL of a solution of 20% piperidine in DMF (200 μL, 2.0 mmol). The solution was stirred for 35 min, when starting material was consumed as confirmed by TLC (4:1 EtOAc:Hexane). Excess acetic anhydride (400 μL, 4.2 mmol, 5 eq.) was added, and the solution was stirred for 16 h. The solution was concentrated *in vacuo* using a secondary trap and

remaining solvent was acidified to pH 2 using 1 M HCl as indicated by pH paper. The aqueous layer was washed with three portions of 20 mL EtOAc, and the combined organic layers were washed with 1 x 10 mL H_2O and brine. Combined organics were dried over Na_2SO_4, filtered, and concentrated *in vacuo*. Flash column chromatography (4:1 EtOAc:Hexane) yielded 170 mg (0.70 mmol, 83.0% yield) as a colorless oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.27 (d, 1H, *J* = 8 Hz, N*H*), 4.37 (m, 1H, C*H*CH_2N_3), 3.48 (m, 2H, CH_2N_3), 2.50 (m, 2H, CH_2(O)O*t*Bu), 2.00 (s, 3H, CH_3(O)NH), 1.46 (s, 9H, -O*t*Bu).



Methyl ester 3.17

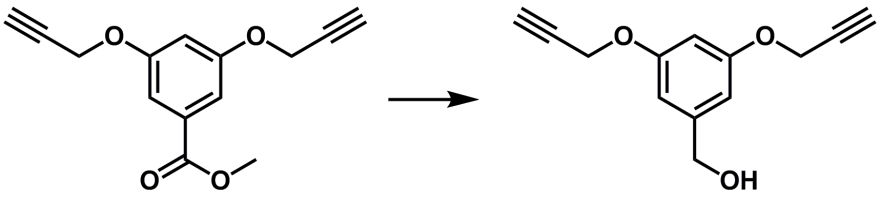
To a stirred solution of methyl-3,5-dihydroxybenzoate (5.00 g, 29.7 mmol, 1 eq.) and propargyl bromide (7.32 mL, 65.4 mmol, 2.2 eq.) in acetone (90 mL) was added K_2CO_3 (4.52 g, 32.7 mmol, 1.1 eq.) and 18-crown-6 (31.4 mg, 0.12 mmol, 0.004 eq.). The solution was heated to reflux and stirred for 18 h. The reaction mixture was filtered, and then evaporated to dryness. The solid residue was re-suspended in methanol and cooled to 0°C, yielding a yellow solid. The solid was filtered, washed with methanol, and dried under reduced pressure to yield 5.624 g yellow solid (23.0 mmol, 77.4% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.30 (d, 2H, *J* = 2.4 Hz, aromatic), 6.81 (t, 1H, *J* = 2.4 Hz, aromatic), 4.72 (d, 4H, *J* = 2.5 Hz, CH_2C≡CH), 3.91 (s, 3H, OCH_3), 2.55 (t, 2H, *J* = 2.5 Hz, CH_2C≡C*H*)



Benzoic Acid 3.8

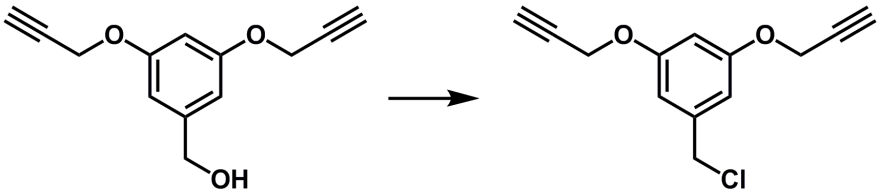
To a stirred solution of methyl ester **21** (500 mg, 2.1 mmol, 1 eq.) in DMF (2 mL) was added aqueous 4M NaOH (2.6 mL, 10 mmol, 5 eq.). The solution was heated to 70°C and stirred until the starting material was consumed as indicated by TLC (3:2 EtOAc:Hexane). The

solution was cooled to room temperature and acidified with 1 M HCl to pH 2, as indicated by pH paper. A yellow solid precipitated from the solution that was collected by vacuum filtration and used without further purification: 424.8 mg, 2.1 mmol, 90.1% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.35 (d, 2H, *J* = 2.4 Hz, aromatic), 6.85 (t, 1H, *J* = 2.4 Hz, aromatic), 4.73 (d, 4H, *J* = 2.4 Hz, C*H*$_2$C≡CH), 2.56 (t, 2H, *J* = 2.5 Hz, CH$_2$C≡C*H*)



Benzyl Alcohol

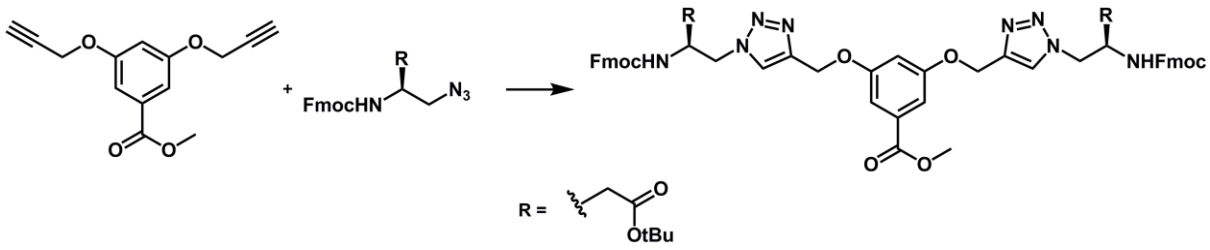
To a stirred suspension of methyl ester **21** (500 mg, 2.0 mmol, 1 eq.) in dry THF (10 mL) was added excess $LiAlH_4$ (715.7 mg, 20.5 mmol, 10 eq.). After stirring at room temperature for 2 h, the solution was cooled to 0°C. Excess $LiAlH_4$ was quenched by subsequent addition of H_2O (720 µL), 3M NaOH (720 µL), and an additional portion of H_2O (2.2 mL). The mixture was stirred at room temperature for 20 min, followed by addition of Na_2SO_4 and stirring for an additional 15 min. The solid was filtered under vacuum and washed with CH_2Cl_2. The filtrate was concentrated under reduced pressure to yield 408 mg solid (1.9 mmol**,** 94.4% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.64 (d, 2H, *J* = 2.4 Hz, aromatic), 6.55 (t, 1H, *J* = 2.4 Hz, aromatic), 4.68 (d, 4H, *J* = 2.4 Hz, C*H*$_2$C≡CH), 4.67 (d, 2H, *J* = 6 Hz, C*H*$_2$OH), 2.53 (t, 2H, *J* = 2.4 Hz, CH$_2$C≡C*H*), 1.66 (t, 1H, *J* = 6 Hz, CH$_2$O*H*)



Benzyl Chloride 3.2

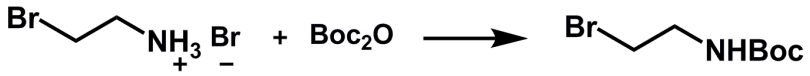
To a cooled (0°C) solution of benzyl alcohol **22** (500 mg, 2.3 mmol, 1 eq.) and dry pyridine (374 µL, 4.6 mmol, 2 eq.) in dry CH_2Cl_2 (5 mL) was added $SOCl_2$ (252 µL, 3.5 mmol, 1.5 eq.) slowly. The solution was allowed to warm to room temperature and was stirred overnight. The reaction was quenched with water (10 mL) and aqueous and organic layers were

separated. The organic layer was washed with 3 portions of 20 mL H_2O. The aqueous layer was washed with two portions of 20 mL CH_2Cl_2. Combined organic layers were dried over Na_2SO_4. The solution was filtered and concentrated *in vacuo* followed by flash column chromatography (2:1 CH_2Cl_2:hexane) to yield 116.8 mg (0.50 mmol, 21.5% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.65 (d, 2H, *J* = 2.4 Hz, aromatic), 6.58 (t, 1H, *J* = 2.3 Hz, aromatic), 4.69 (d, 4H, *J* = 2.4 Hz, C*H*$_2$C≡CH), 4.53 (s, 2H, C*H*$_2$Cl), 2.54 (t, 2H, *J* = 2.4 Hz, CH$_2$C≡C*H*)



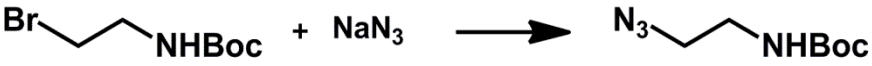
Click Dendron

To a stirred solution of azide (573 mg, 1.4 mmol, 2 eq.) in 1:1:1 *t*-BuOH/H_2O/CH_2Cl_2 (6 mL) was added alkyne (166 mg, 0.7 mmol, 1 eq.). To this solution was added Na ascorbate (13.4 mg, 0.07 mmol, 0.1 eq.) and $CuSO_4$-5H_2O (16.9 mg, 0.07 mmol, 0.1 eq.). The reaction was stirred at room temperature under N_2 overnight. The solution was diluted with 10 mL H_2O and 1 mL concentrated NH_4OH and stirred for ten minutes. It was then transferred to a separatory funnel and extracted with three 20 mL portions of EtOAc. The combined organic layers were washed with one 10 mL portion of saturated NaCl, dried over Na_2SO_4, and concentrated *in vacuo*. Flash column chromatography (3:1 EtOAc:Hexane) yielded the desired product as a white solid (452 mg, 0.4 mmol, 61% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.72 (d, 4H, *J* = 7.5 Hz, Fmoc ar *H*), 7.56 (s, 2H, triazole C=C*H*N), 7.54 (t, 4H, *J* = 3.6 Hz, Fmoc ar *H*), 7.37 (t, 4H, *J* = 7.5 Hz, Fmoc ar *H*), 7.29 (t, 4H, *J* = 7.5 Hz, Fmoc ar *H*), 7.27 (s, 2H, central ar *H*), 6.77 (s, 1H, central ar *H*), 5.65 (d, 2H, *J* = 8 Hz, CHN*H*C(O)), 5.13 (s, 4H, OC*H*$_2$C(N)=C), 4.60 (m, 4H, *J* = 7.2 Hz, Fmoc C*H*$_2$), 4.39 (m, 6H, NHC*H*C*H*$_2$N [peak overlap]), 4.17 (t, 2p, *J* = 6.7 Hz, Fmoc C*H*), 3.88 (s, 3H, C(O)OC*H*$_3$), 2.47 (d, 4H, *J* = 5.7 Hz, CHC*H*$_2$C(O)), 1.44 (s, 18H, OC(C*H*$_3$)$_3$). LR-MS (ESI): calc. (M+H): 1089.2. Found (M+H): 1089.9.



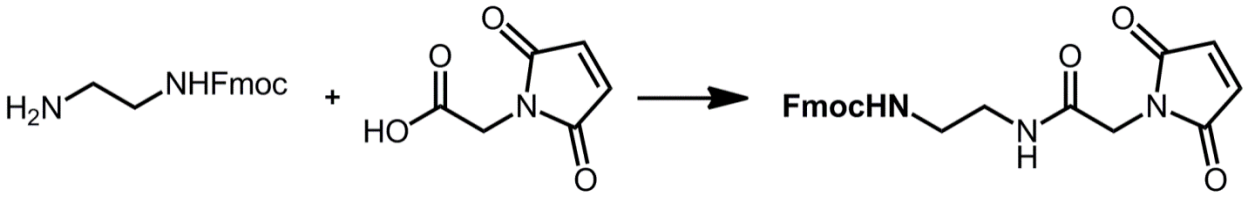
Boc-protected 2-bromoethylamine 3.22

To a stirred solution of 2-bromoethylamine (3.0 g, 14.6 mmol, 1 eq.) in dry CH_2Cl_2 (20 mL) was added Boc_2O (3.2 g, 14.6 mmol, 1 eq.), and solution was cooled to 0°C with an ice bath. Triethylamine (2.5 mL, 17.6 mmol, 1.2 eq.) was added to the solution dropwise over 20 min. The solution was stirred at 0°C for 30 min, warmed to room temperature, and stirred for 24 hr. Reaction mixture was diluted with CH_2Cl_2 (10 mL), transferred to a separatory funnel, and washed with 20 mL portions of 1M $KHSO_4$, H_2O, and brine. The organic layer was dried over Na_2SO_4, filtered, and concentrated by rotary evaporation to yield the Boc-protected 2-bromoethylamine as a clear oil (2.7 g, 11.2 mmol, 83% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 4.95 (bs, 1H, $CH_2NHC(O)$), 3.53 (t, 2H, J = 5.5 Hz, CH_2CH_2NH), 3.45 (t, 2H, J = 5.5 Hz, $BrCH_2CH_2$), 1.45 (s, 9H, $OC(CH_3)_3$)



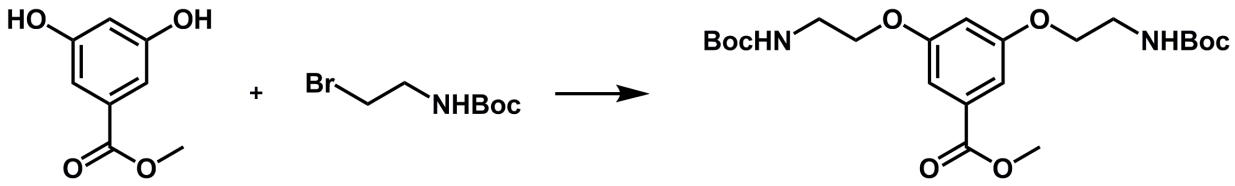
Boc-protected 2-azidoethylamine 3.27

To a stirred solution of Boc-protected 2-bromoethylamine (500.0 mg, 2.2 mmol, 1 eq.) in dry DMF (10 mL) was added NaN_3 (725.2 mg, 11.2 mmol, 5 eq.). Solution was stirred at room temperature under N_2 overnight. Following stirring, solvent was removed *in vacuo* and resuspended in EtOAc (30 mL). Solution was transferred to a separatory funnel and washed twice with H_2O (20 mL) and once with brine (20 mL). The organic layer was dried over Na_2SO_4, filtered into a tared round bottom flask, and concentrated by rotary evaporation to yield the product as a yellow oil (373.3 mg, 89.9% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ 4.81 (bs, 1H, $CH_2NHC(O)$), 3.42 (t, 2H, J = 5.5 Hz, CH_2CH_2NH), 3.31 (t, 2H, J = 5.5 Hz, $N_3CH_2CH_2$), 1.45 (s, 9H, $OC(CH_3)_3$). ESI-MS calc. (M+H): 187.2. Found (M+H): 187.3.



Fmoc-Protected Maleimide Linker 3.26

To a stirred solution of glycine-maleimide (84.2 mg, 0.5 mmol, 1.5 eq.) in dry DMF (5 mL) was added PyBOP (368.6 mg, 0.7 mmol, 2 eq.), HOBt-H_2O (108.5 mg, 0.7 mmol, 2 eq.), and DIPEA (184.8 µL, 1.1 mmol, 3 eq.). Solution was stirred at room temperature for 10 minutes. To this solution was added Fmoc-protected ethylene diamine (100.0 mg, 0.4 mmol, 1 eq.). Solution was stirred at room temperature overnight. Following stirred, solvent was removed *in vacuo* and solid was resuspended in 30 mL EtOAc. Solution was transferred to a separatory funnel and washed with H_2O (20 mL), 1M $KHSO_4$ (20 mL), saturated $NaHCO_3$ (20 mL), and brine (20 mL). The organic layer was dried over Na_2SO_4 and concentrated under reduced pressure. Purification using flash column chromatography (95:5 CH_2Cl_2:MeOH) yielded the product as a clear oil (17.4 mg, 11.7% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.77 (d, 2H, *J* = 7.6 Hz, Fmoc ar), 7.59 (2, 2H, *J* = 7.2 Hz, Fmoc ar), 7.41 (t, 2H, *J* = 7.6 Hz, Fmoc ar), 7.32 (t, 2H, *J* = 7.6 Hz, Fmoc ar), 6.75 (s, 2H, CC*H*=C*H*C), 6.41 (bs, 1H, N*H*), 5.24 (bs, 1H, N*H*), 4.39 (d, 2H, *J* = 6.8 Hz, Fmoc CH_2), 4.21 (t, 1H, *J* = 6.8 Hz, Fmoc C*H*), 4.16 (s, 2H, C(O)CH_2N), 3.37 (t, 2H, *J* = 6 Hz, FmocNHCH$_2$CH_2), 3.32 (t, 2H, *J* = 6 Hz, FmocNHCH_2CH$_2$). ESI-MS calc. (M+H): 420.4. Found (M+H): 420.2.



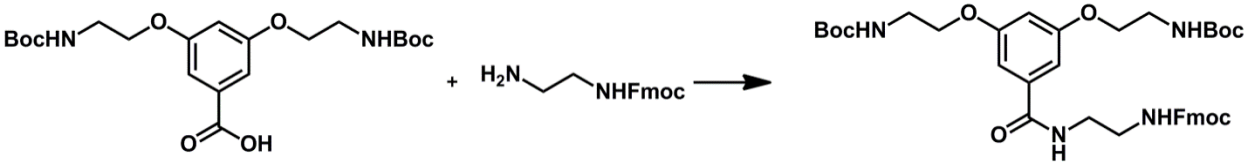
Boc-Protected Methyl ester 3.23

To a stirred solution of Boc-protected (2-bromoethyl) amine (1.5 g, 6.7 mmol, 2.5 eq.) in dry DMF (10 mL) was added 3,5-dihydroxymethyl benzoate (0.45 g, 2.7 mmol, 1 eq.). To this solution was added K_2CO_3 (1.7 g, 12.0 mmol, 4.5 eq.) and 18-crown-6 (2.8 mg, 0.01 mmol, 0.004 eq.). Solution was heated to 40°C and stirred under nitrogen gas for 16 hr. Solution was filtered through a medium frit Buchner funnel to remove a brown solid, and concentrated *in vacuo*. The residue was redissolved in EtOAc (30 mL), transferred to a separatory funnel and washed with two 20 mL portions of H_2O and 20 mL brine. Organic layer was dried over Na_2SO_4, filtered, and concentrated *in vacuo*, yielding the desired product as a clear oil (970 mg, 80% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.17 (d, 2H, *J* = 2 Hz, ar *H*), 6.63 (t, 1H, *J* = 2 Hz, ar

H), 4.98 (bs, 2H, $CH_2NHC(O)$), 4.03 (t, 4H, *J* = 5 Hz, OCH_2CH_2NH), 3.89 (s, 3H, $C(O)OCH_3$), 3.53 (q, 4H, *J* = 5 Hz, OCH_2CH_2NH), 1.45 (s, 18H, $OC(CH_3)_3$)

Boc-Protected Acid

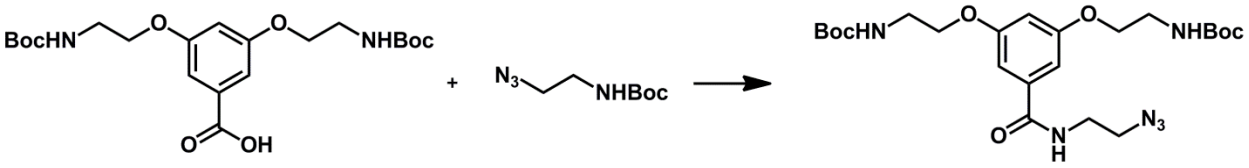
To a stirred solution of Boc-protected methyl ester (478.5 mg, 1.1 mmol, 1 eq.) in 14:5 dioxane:MeOH (19 mL) was added NaOH (210.6 mg, 5.3 mmol, 5 eq.). Solution was stirred at room temperature overnight. Following stirring, the reaction was acidified to pH 2 using 1M HCl. Solvent was removed *in vacuo* and solid was resuspended in 30 mL EtOAc and transferred to a separatory funnel. Solvent was washed twice with 20 mL 1M HCl, twice with 20 mL 1M $KHSO_4$, and twice with 20 mL brine. Organic layer was dried over Na_2SO_4, filtered, and concentrated *in vacuo*, yielding the desired product as a clear oil (388.5 mg, 83.8%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.19 (d, 2H, *J* = 4 Hz, ar *H*), 6.64 (t, 1H, *J* = 4 Hz, ar *H*), 5.06 (bs, 2H, $CH_2NHC(O)$), 4.03 (t, 4H, *J* = 5 Hz, OCH_2CH_2NH), 3.54 (q, 4H, *J* = 5 Hz, OCH_2CH_2NH), 1.45 (s, 18H, $OC(CH_3)_3$). ESI-MS (M+H): calc.: 441.2. Found (M+H): 441.2.



Boc-Protected Fmoc-Amine Core 3.24

To a stirred solution of Boc-protected acid (400.9 mg, 0.9 mmol, 1 eq.) in dry DMF (5 mL) was added PyBOP (710.8 mg, 1.4 mmol, 1.5 eq.), HOBt-H_2O (209.2 mg, 1.4 mmol, 1.5 eq.), and DIPEA (475.2 µL, 2.7 mmol, 3 eq.). Solution was allowed to stir at room temperature for 10 minutes. To this solution was added Fmoc-protected ethylene diamine (385.6 mg, 1.4 mmol, 1.5 eq.). Solution was allowed to stir at room temperature overnight. Following stirring, solvent was removed *in vacuo* and solid was resuspended in 30 mL EtOAc and transferred to a separatory funnel. The organic layer was washed with 20 mL H_2O, 20 mL 6M HCl, 20 mL H_2O,

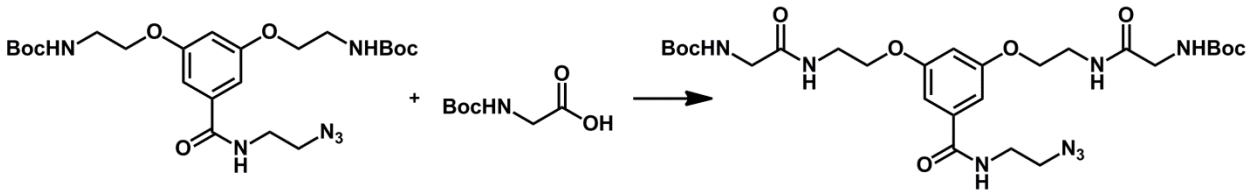
20 mL 4M NaOH, 20 mL H_2O, and 20 mL brine. Solution was dried over Na_2SO_4, filtered, and concentrated by rotary evaporation. Purification by flash column chromatography (2:1 EtOAc:Hexane) yielded the product as a white solid (193.8 mg, 30% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.74 (d, 2H, *J* = 7.6 Hz, Fmoc ar), 7.55 (2, 2H, *J* = 7.2 Hz, Fmoc ar), 7.38 (t, 2H, *J* = 7.2 Hz, Fmoc ar), 7.27 (t, 2H, *J* = 7.2 Hz, Fmoc ar), 7.03 (bs, 1H, FmocN*H*) 6.94 (d, 2H, *J* = 4 Hz, ar *H*), 6.50 (t, 1H, *J* = 4 Hz, ar *H*), 5.39 (bs, 1H, ArC(O)N*H*) 4.97 (bs, 2H, CH_2NHC(O)), 4.40 (d, 2H, *J* = 6.8 Hz, Fmoc CH_2), 4.17 (t, 1H, *J* = 6 Hz, Fmoc C*H*) 3.96 (t, 4H, *J* = 5 Hz, OCH_2CH_2NH), 3.58 (m, 2H, FmocNHCH$_2$CH_2), 3.47 (m, 5H, OCH$_2$CH_2NH and FmocNHCH_2), 1.44 (s, 18H, OC(CH_3)$_3$). ESI-MS (M+H) calc.: 705.8. Found (M+H): 705.3. Elemental Analysis calc. CHN: 64.7:6.9:8.0. Found CHN: 63.1:6.7:7.4.



Boc-Protected Azide Core 3.28

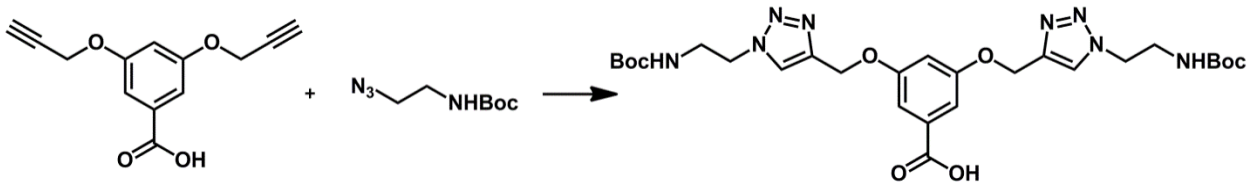
To a stirred solution of Boc-protected 2-azidoethylamine (85.9 mg, 0.5 mmol, 1.5 eq.) in dry CH_2Cl_2 (5 mL) was added TFA (5 mL). Solution was stirred at room temperature for 1 hr. After stirring, solvent was removed *in vacuo* and stored under vacuum overnight. In a second vial, to a stirred solution of Boc-protected acid (135.4 mg, 0.3 mmol, 1 eq.) was added PyBOP (240.0 mg, 0.5 mmol, 1.5 eq.), HOBt-H_2O (70.6 mg, 0.5 mmol, 1.5 eq.), and DIPEA (160.5 µL, 0.9 mmol, 3 eq.). Solution was stirred at room temperature for 10 minutes. Solution was transferred to vial containing 2-azidoethylamine and stirred at room temperature overnight. Following stirring, solvent was removed by azeotropic distillation and solid was resuspended in 30 mL EtOAc and transferred to a separatory funnel. The organic layer was washed with 20 mL H_2O, 20 mL 6M HCl, 20 mL H_2O, 20 mL 4M NaOH, 20 mL H_2O, and 20 mL brine. Solution was dried over Na_2SO_4, filtered, and concentrated by rotary evaporation. Purification by flash column chromatography (3:1 EtOAc:Hexane) yielded the product as a white solid (83.1 mg, 53.2% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.97 (bs, 1H, N_3CH2CH2N*H*), 6.93 (d, 2H, *J* = 1.2 Hz, ar *H*), 6.48 (t, 1H, *J* = 2.4 Hz, ar *H*), 5.15 (bs, 2H, CH_2NHC(O)), 3.97 (t, 4H, *J* = 5.2 Hz, OCH_2CH_2NH), 3.60 (m, 2H, $N_3CH_2CH_2$), 3.54 (m, 2H, $N_3CH_2CH_2$), 3.48 (m, 4H,

OCH_2CH_2NH), 1.44 (s, 18H, $OC(CH_3)_3$). ESI-MS (M+H) calc.: 509.3. Found (M+H): 509.4. Elemental Analysis calc. CHN: 54.32:7.13:16.52. Found CHN: 54.24:7.01:15.32.



Extended Boc-Protected Azide Core 3.30

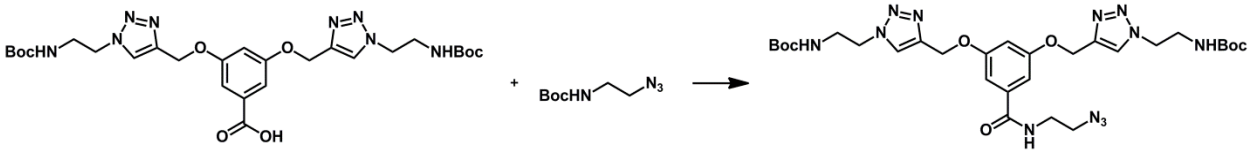
To a stirred solution of Boc-protected azide core (180.8 mg, 0.4 mmol, 1 eq.) in CH_2Cl_2 (3 mL) was added TFA. Solution was stirred at room temperature for 1 hr. Following stirring, solvent was removed *in vacuo* and remaining solid was stored under vacuum. In a separate flask, to a stirred solution of Boc-protected glycine (187.0 mg, 1.1 mmol, 3 eq.) was added PyBOP (555.2 mg, 1.1 mmol, 3 eq.), HOBt-H_2O (163.4 mg, 1.1 mmol, 3 eq.), and DIPEA (371.2 µL, 2.1 mmol, 6 eq.). Solution was stirred at room temperature for 10 minutes, then transferred to flask containing the azide core and stirred at room temperature overnight. Following stirring, solvent was removed by azeotropic distillation and solid was resuspended in 30 mL EtOAc and transferred to a separatory funnel. The organic layer was washed with 20 mL H_2O, 20 mL 6M HCl, 20 mL H_2O, 20 mL 4M NaOH, 20 mL H_2O, and 20 mL brine. Solution was dried over Na_2SO_4, filtered, and concentrated by rotary evaporation. Product was identified by ESI-MS (M+H) calc.: 623.7. Found (M+H): 623.4.



Click Core Acid

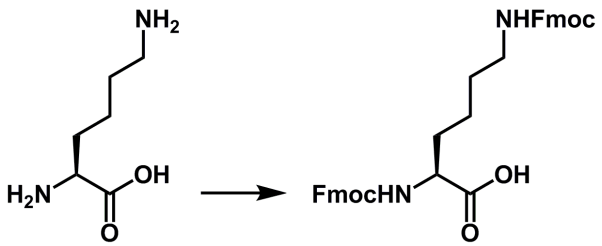
To a stirred solution of Boc-protected 2-azidoethylamine (250.0 mg, 1.3 mmol, 2.1 eq.) in 1:1 THF:H_2O (6 mL) was added benzoic acid (147.2 mg, 0.6 mmol, 1 eq.). To this solution was added sodium ascorbate (63.3 mg, 0.3 mmol, 0.5 eq.) and $CuSO_4$-$5H_2O$ (79.8 mg, 0.3 mmol, 0.5 eq.). Solution was stirred at room temperature overnight. Following stirring, solvent was removed *in vacuo* and solid was resuspended in 30 mL saturated $NaHCO_3$. Aqueous layer was

extracted with three portions of 20 mL CH_2Cl_2. A white solid crashed out from extraction solution that was collected by vacuum filtration. Organic layer was dried over Na_2SO_4, filtered, and concentrated *in vacuo* to yield a small amount of product. Collected white solid was identified by ESI-MS as desired product (384.3 mg, 99.8% yield). LR-MS (ESI): calc. (M+Na): 625.6. Found (M+Na): 625.4.



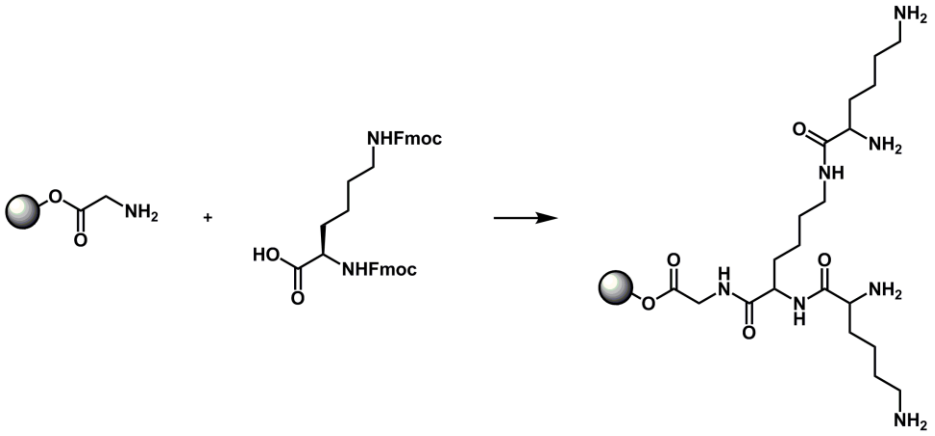
Click Core Azide 3.31

To a stirred solution of Boc-protected 2-azidoethylamine (149.4 mg, 0.8 mmol, 1.5 eq.) in CH_2Cl_2 (3 mL) was added TFA (3 mL). Solution was stirred at room temperature for 1 hr. Following stirring, solvent was removed *in vacuo* and solid was stored under vacuum overnight. In a separate flask, to a stirred solution of click core acid (322.2 mg, 0.5 mmol, 1 eq.) in dry DMF (5 mL) was added PyBOP (417.5 mg, 0.8 mmol, 1.5 eq.) and DIPEA (279.1 µL, 1.6 mmol, 3 eq.). Solution was stirred at room temperature for 1 hr. Solution was then transferred to flask containing 2-azidoethylamine and stirred at room temperature overnight. Following stirring, solvent was removed by azeotropic distillation and solid resuspended in 30 mL saturated $NaHCO_3$. Aqueous layer was extracted with three portions of 20 mL CH_2Cl_2. Combined organic layers were dried over Na_2SO_4, filtered, and concentrated by rotary evaporation. Purification by flash column chromatography (95:5 CH_2Cl_2:MeOH) yielded the desired product as identified by ESI-MS. LR-MS (ESI): calc. (M+H): 671.7. Found (M+H): 671.3.



Fmoc-protected Lysine

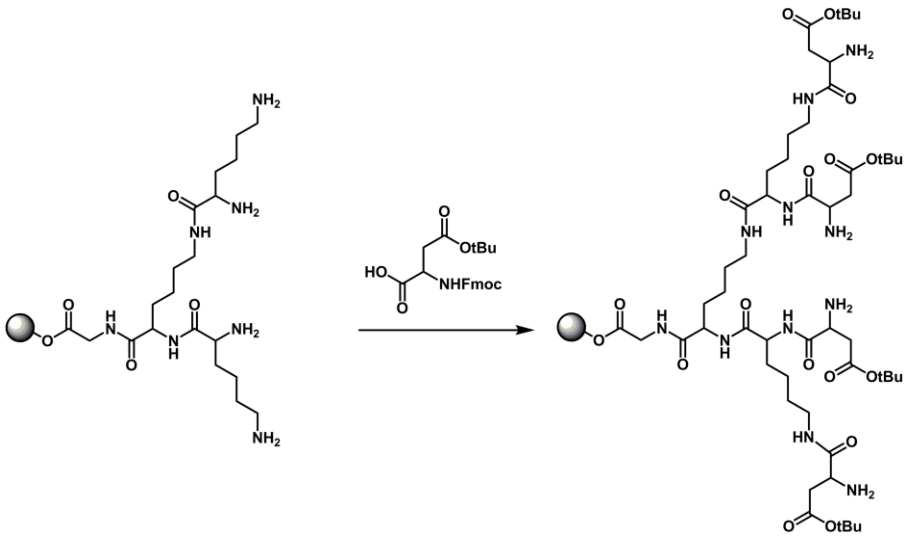
Lysine-HCl (1.0 g, 5.5 mmol, 1 eq.) was dissolved in H_2O (200 mL) and 1,4-dioxane (80 mL). To this stirred solution was added $NaHCO_3$ (965.9 mg, 11.5 mmol, 2.1 eq.), followed by Fmoc-OSuc (3.7 g, 11.0 mmol, 2 eq.) as a solution in 1,4-dioxane dropwise. The solution was stirred overnight under air at room temperature. Addition of H_2O (400 mL) and acidification to pH 1 with 1M HCl yielded a white solid. The solid was filtered through a 150 mL course frit filter and dried under vacuum over Dryerite overnight, yielding the protected lysine as a white solid (2.8 g, 4.7 mmol, 86% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.74 (d, 4H, *J* = 7 Hz, Fmoc ar *H*), 7.56 (t, 4H, *J* = 6.3 Hz, Fmoc ar *H*), 7.37 (t, 4H, *J* = 7.2 Hz, Fmoc ar *H*), 7.28 (t, 4H, *J* = 7.4 Hz, Fmoc ar *H*), 5.60 (m, 2H, OC(O)N*H*), 4.87 (m, 1H, NHC*H*C(O)), 4.38 (m, 4H, Fmoc C*H*$_2$), 4.17 (t, 2H, *J* = 6.2 Hz), 3.19 (m, 2H, *J* = 7 Hz, $CHCH_2CH_2CH_2CH_2$NHC(O)), 1.77 (m, 2H, $CHCH_2CH_2CH_2CH_2$NHC(O)), 1.52 (m, 2H, $CHCH_2CH_2CH_2CH_2$NHC(O)), 1.40 (m, 2H, $CHCH_2CH_2CH_2CH_2$NHC(O)). LR-MS (ESI): calc. (M+H): 591.7. Found (M+H): 591.5.



Lysine Dendron 3.19 – Standard Synthesis

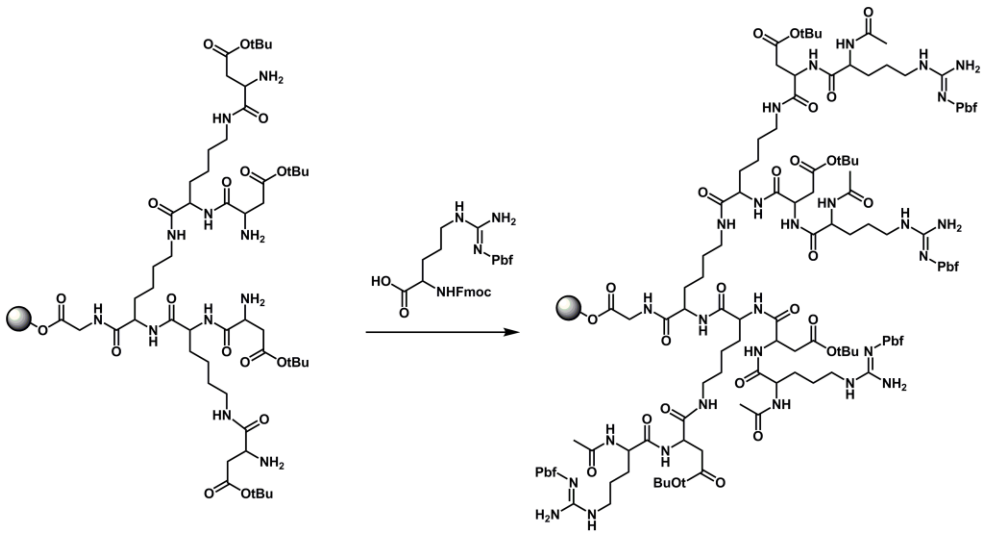
To a stirred solution of dried Fmoc-Lys(Fmoc)-OH (42.5 mg, 0.07 mmol, 3 eq.) in dry DMF (1 mL) was added PyBOP (37.5 mg, 0.07 mmol, 3 eq.), HOBt-H_2O (11.0 mg, 0.07 mmol, 3 eq.), and DIPEA (16.7 μL, 0.1 mmol, 4 eq.). This solution was transferred to a solid phase peptide synthesis (SPPS) vessel containing pre-swelled 2-chlorotrityl resin loaded with glycine (30.0 mg, 0.02 mmol, 1 eq.). The resin and peptide/coupling mixture were gently mixed with N_2 gas bubbling through the vessel. The reaction was monitored by ninhydrin test[21] and MALDI-MS. The vessel was drained by vacuum filtration, and beads rinsed with 2 mL DMF three times. Fmoc deprotection was accomplished by the addition of 2 mL 20% piperidine in DMF and mixing with N_2 gas three times, washing the beads with DMF between additions. The coupling

and deprotection steps were repeated with twice the equivalents of coupling reagents, yielding the lysine dendron which was used without further purification. LR-MS containing Fmoc-protected amines (MALDI): calc. (M+H) 1348.5. Found (M+H): 1348.8.



Lysine-Asp-NH_2 Dendron 3.20 – Standard Synthesis

To a stirred solution of dried Fmoc-Asp(OtBu)-OH (118.5 mg, 0.3 mmol, 12 eq.) in dry DMF (2 mL) was added PyBOP (149.9 mg, 0.3 mmol, 12 eq.), HOBt-H_2O (44.1 mg, 0.3 mmol, 12 eq.), and DIPEA (66.8 μL, 0.4 mmol, 16 eq.). This solution was transferred to a solid phase peptide synthesis (SPPS) vessel containing pre-swelled 2-chlorotrityl resin loaded with the Lys-NH_2 dendron (30.0 mg, 0.02 mmol, 1 eq.). The resin and peptide/coupling mixture were gently mixed with N_2 gas bubbling through the vessel overnight. The reaction was monitored by MALDI-MS**.** The vessel was drained by vacuum filtration, and beads rinsed with 2 mL DMF three times. Fmoc deprotection was accomplished by the addition of 2 mL 20% piperidine in DMF and mixing with N_2 gas three times, washing the beads with DMF between additions, which yielded the dendron without further purification. LR-MS containing Fmoc-protected amines (MALDI): calc. (M+H) 2033.3. Found (M+H): 2035.2.



Lysine-Asp-Arg-Ac Dendron 3.21 – Standard Synthesis

To a stirred solution of dried Fmoc-Arg(Pbf)-OH (186.9 mg, 0.3 mmol, 12 eq.) in dry DMF (2 mL) was added PyBOP (149.9 mg, 0.3 mmol, 12 eq.), HOBt-H_2O (44.1 mg, 0.3 mmol, 12 eq.), and DIPEA (66.8 μL, 0.4 mmol, 16 eq.). This solution was transferred to a solid phase peptide synthesis (SPPS) vessel containing pre-swelled 2-chlorotrityl resin loaded with the Lys-Asp-NH_2 dendron (30.0 mg, 0.02 mmol, 1 eq.). The resin and peptide/coupling mixture were gently mixed with N_2 gas bubbling through the vessel overnight. The reaction was monitored by MALDI-MS**.** The vessel was drained by vacuum filtration, and beads rinsed with 2 mL DMF three times. Fmoc deprotection was accomplished by the addition of 2 mL 20% piperidine in DMF and mixing with N_2 gas three times, washing the beads with DMF between additions. After deprotection, added 2 mL acetic anhydride and mixed with N_2 gas bubbling through the vessel overnight. The vessel was drained and rinsed as described above, yielding the acyl-protected dendron. LR-MS containing Fmoc-protected amines (MALDI): calc. (M+H) 3667.4. Found (M+H): 3671.3. LR-MS containing Acyl-protected amines (MALDI): calc. (M+H) 2946.6. Found (M+H): 2950.1.

Dendron Derivative Characterization

For each completed dendron, a small sample of resin was transferred to a glass vial with 1 mL 20% HFIP in CH_2Cl_2 and allowed to stand for 1 hr. Samples were then filtered through a glass-wool plugged pipette into a second glass vial, concentrated *in vacuo*, and submitted for

low-res MALDI-MS. The following are completed dendrons containing side-chain protecting groups, with peripheral dipeptides and N-terminal protecting group indicated.

Gly-DR-acetyl: LR-MS: calc. (M+H): 2947.6, Found (M+H): 2947.3

Gly_2-DR-acetyl: LR-MS: calc. (M+H): 3004.6, Found (M+H): 3004.5

Gly_3-DR-acetyl: LR-MS: calc. (M+H): 3061.7, Found (M+H): 3061.8

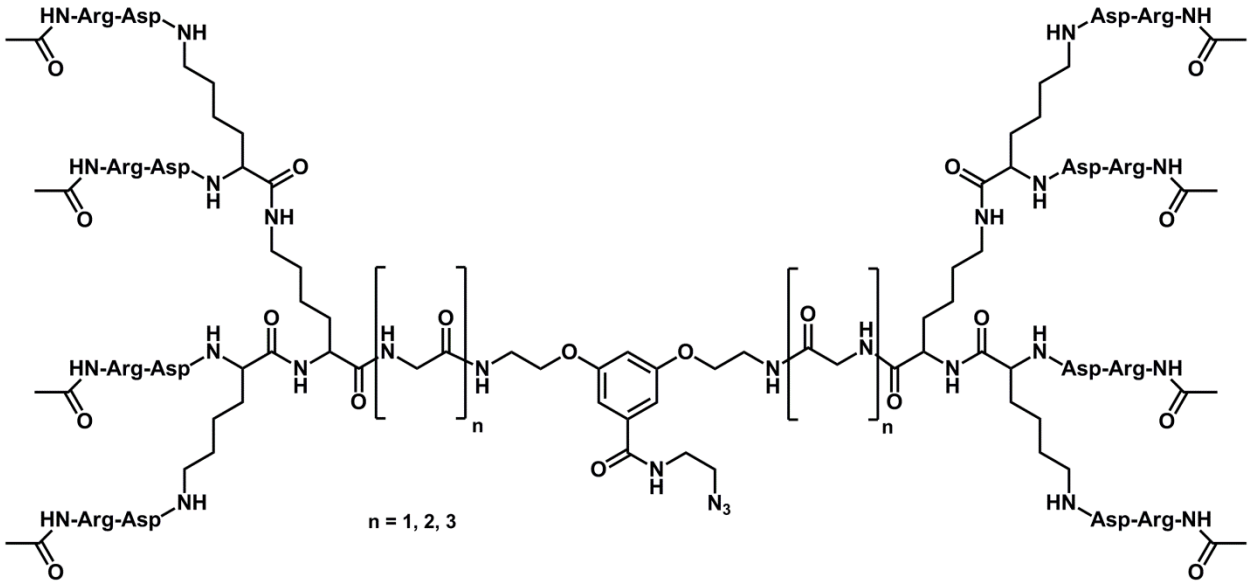
Gly_3-$(DR)_2$-acetyl: LR-MS: calc. (M+H minus Pbf protecting group): 5128.2, Found (M+H minus Pbf protecting group): 5129.6

Gly_3-D_2R_2-acetyl: LR-MS: calc. (M+H): 5380.5, Found (M+H): 5380.3

Gly_3-DA-acetyl: LR-MS: calc. (M+Na): 1733.9, Found (M+Na): 1732.5

Gly_3-AR-acetyl: LR-MS: calc. (M+H): 2659.3, Found (M+H): 2659.3

Gly_3-AA-Fmoc: LR-MS: calc. (M+Na): 2052.9, Found (M+Na): 2052.8



Full Dendrimer Synthesis – Boc-Protected Aromatic Core + Peptide Dendron (3.29)

To a stirred solution of Boc-protected aromatic core containing azide linker (17.2 mg, 0.03 mmol, 1 eq.) in CH_2Cl_2 (3 mL) was added TFA (3 mL). Solution was stirred at room temperature for 1 hr. Following stirring, solvent was removed *in vacuo* and stored under vacuum. In a separate flask, to a stirred solution of peptide dendron (398.2 mg, 0.1 mmol, 4 eq.) in dry DMF (3 mL) was added PyBOP (140.6 mg, 0.3 mmol, 8 eq.), HOBt-H_2O (41.4 mg, 0.3 mmol, 8 eq.), and DIPEA (47.0 µL, 0.3 mmol, 8 eq.). Solution was stirred at room temperature for 10 minutes, then transferred to flask containing aromatic core and stirred at room temperature under N_2 overnight. Following stirring, solvent was removed by azeotropic distillation and solid was resuspended in 30 mL EtOAc and transferred to a separatory funnel. Organic layer was washed with 20 mL H_2O, 20 mL 6M HCl, 20 mL H_2O, 20 mL 4M NaOH, 20 mL H_2O, and 20 mL brine. Organic layer was dried over Na_2SO_4, filtered, and concentrated by rotary evaporation. A small amount of desired product was identified by MALDI-MS, with a majority being the single-addition product.

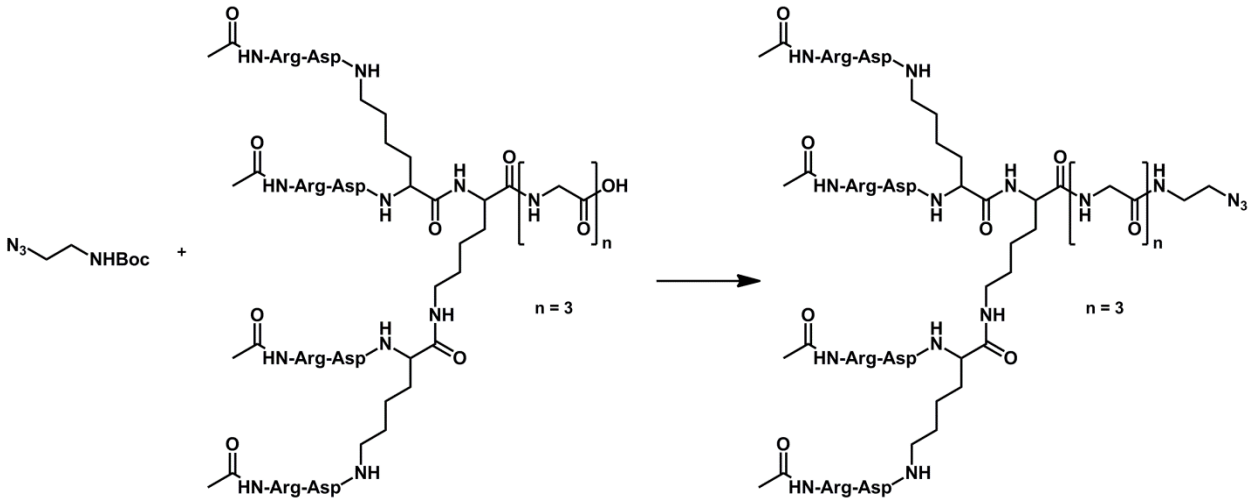
Full Dendrimer Characterization

Products listed describe composition of peptide dendron side chain. Single-armed and full dendrimers are listed when found.

Gly-DR-acetyl – single-arm: LR-MS: calc. (M+H): 3237.9, Found (M+H): 3238.0

Gly-DR-acetyl – full: LR-MS: calc. (M+H): 6166.4, Found (M+H): 6166.5

Gly_2-DR-acetyl – single-arm: LR-MS: calc. (M+H): 3294.9, Found (M+H): 3295.5



Dendron + 2-azidoethylamine 3.32

To a stirred solution of Boc-protected 2-azidoethylamine (33.3 mg, 0.2 mmol, 5 eq.) in CH_2Cl_2 (2 mL) was added trifluoroacetic acid (2 mL). Solution was stirred at room temperature under N_2 for 1 hr. Solvent was removed *in vacuo* and remaining solid was stored under vacuum overnight to remove excess TFA. In a separate vial, to a stirred solution of dendron (109.4 mg, 0.04 mmol, 1 eq.) in dry DMF (3 mL) was added PyBOP (93.0 mg, 0.2 mmol, 5 eq.), HOBt-H_2O (27.4 mg, 0.2 mmol, 5 eq.), and DIPEA (31.1 µL, 0.2 mmol, 5 eq.). Solution was stirred at room temperature for 1 hr. Solution was then transferred to flask containing 2-azidoethylamine and stirred at room temperature under N_2 overnight. Solvent was removed by azeotropic distillation, and solid was resuspended in 30 mL EtOAc and transferred to a separatory funnel. The organic layer was washed with 20 mL H_2O, 20 mL 1M HCl, 20 mL H_2O, 20 mL 4M NaOH, 20 mL H_2O, and 20 mL brine. The organic layer was dried over Na_2SO_4, filtered into a tared round bottom flask, and concentrated by rotary evaporation to yield the product as an off-white solid (31.5 mg, 28.2% yield).

Dendron-Azide Linker Derivative Characterization

Products were characterized by MALDI-MS. The following are completed dendrons containing azide linker and side-chain protecting groups, with peripheral dipeptides and N-terminal protecting group indicated.

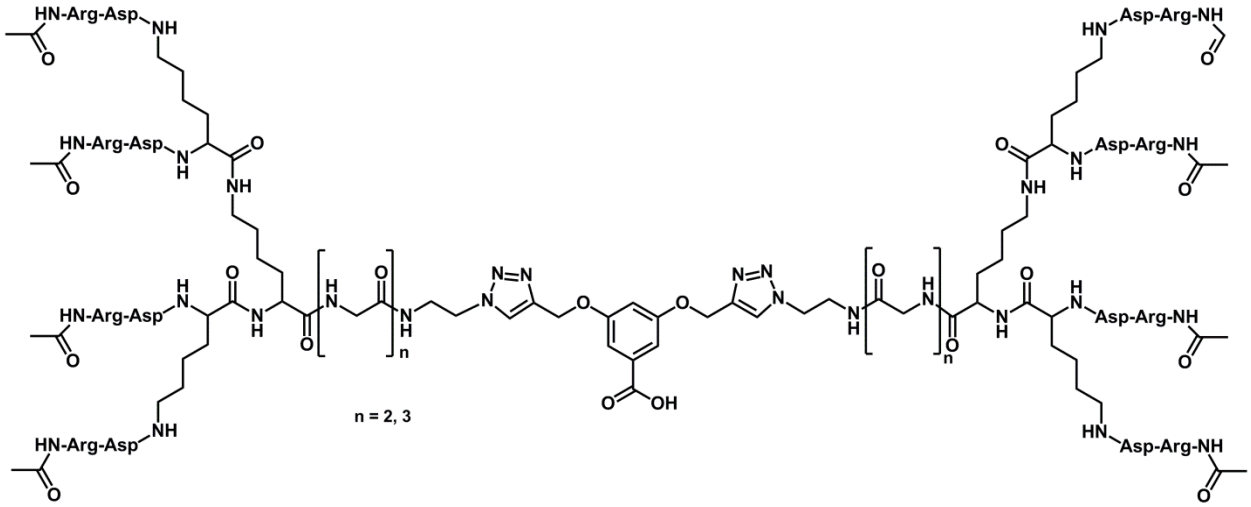
<u>Gly_2-DR-acetyl</u>: LR-MS: calc. (M+H): 3071.7, Found (M+H): 3071.4

$\underline{Gly_3\text{-DR-acetyl}}$: LR-MS: calc. (M+H): 3128.8, Found (M+H): 3128.9

$\underline{Gly_3\text{-}(DR)_2\text{-acetyl}}$: LR-MS: calc. (M+H): 5447.6, Found (M+H): 5447.6

$\underline{Gly_3\text{-AR-acetyl}}$: LR-MS: calc. (M+H minus Pbf protecting group): 2476.0, Found (M+H minus Pbf protecting group): 2476.0

$\underline{Gly_3\text{-AA-acetyl}}$: LR-MS: calc. (M+H): 1378.5, Found (M+H): 1387.8



Full Dendrimer – Copper Catalyzed Click Reaction with Acid Core

To a stirred solution of peptide dendron containing azide linker (53.4 mg, 0.02 mmol, 2 eq.) in 1:1 THF:H_2O (2 mL) was added aromatic core (2.0 mg, 0.01 mmol, 1 eq.), sodium ascorbate (0.9 mg, 0.004 mmol, 0.5 eq.), and $CuSO_4$-$5H_2O$ (0.1 mg, 0.004 mmol, 0.05 eq.). Solution was heated to 100°C for 10 minutes in a microwave reactor. A solid crashed out of solution, which was washed with three portions of 1 mL MeOH. MALDI-MS indicated that the single-armed product was isolated.

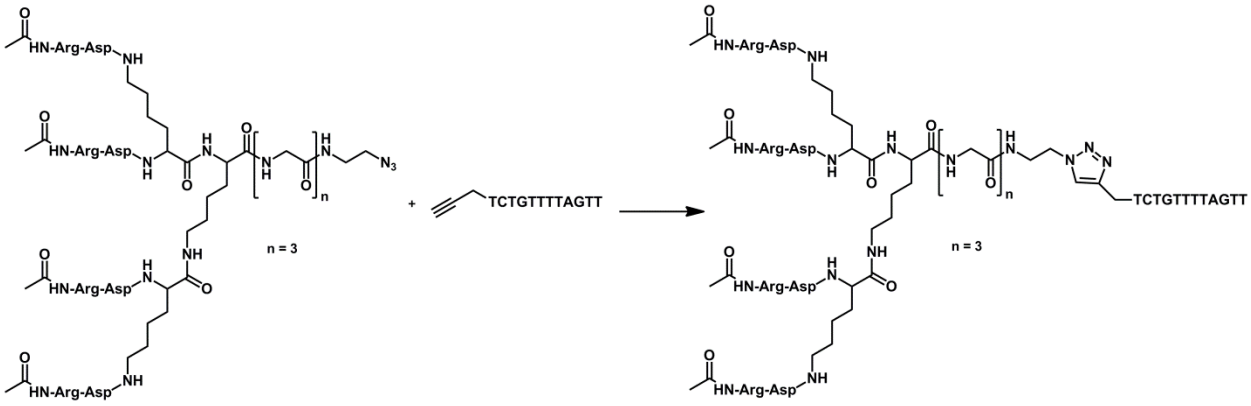
Full Dendrimer Characterization

Products listed describe composition of peptide dendron side chain. Single-armed and full dendrimers are listed when found.

$\underline{Gly_2\text{-DR-acetyl – single-arm, protected side-chains}}$: LR-MS: calc. (M+H): 3301.9, Found (M+H): 3308.6

Gly_2-DR-acetyl – single-arm: LR-MS: calc. (M+H): 2068.2, Found (M+H): 2069.3

Gly_3-DR-acetyl – single-arm, protected side-chains: LR-MS: calc. (M+H): 3358.5, Found (M+H): 3361.5



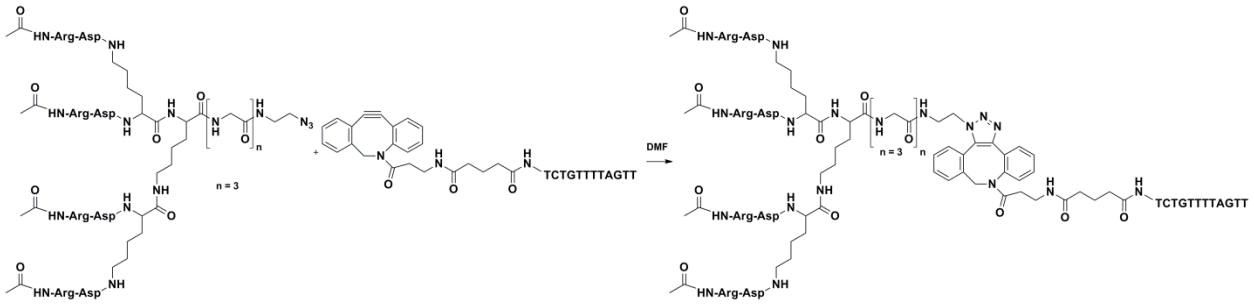
Copper Catalyzed Click Dendrimer Conjugation 3.33

To a solution of peptide dendron Gly_3-DR-acetyl containing azide linker (1.9 mg, 1.0 µmol, 10 eq.) in CH_2Cl_2 (2 mL) was added TFA (2 mL) and solution was stirred at room temperature for 1 hr. Solvent was removed *in vacuo* and solid was stored under vacuum. Solid was resuspended in 200 µL DMF. To this solution was added SMN antisense oligonucleotide containing 3'-alkyne modifier (250 µL, 0.1 µmol, 1 eq.), 1M sodium ascorbate (4 µL, 4 µmol, 40 eq.), and 1M $CuSO_4$-$5H_2O$ (4 µL, 4 µmol, 40 eq.). Solution was heated in a thermal cycler at 94°C for 30 minutes, then 72°C for 30 minutes. Product was analyzed by MALDI-MS, indicating that desired product had been synthesized. LR-MS: calc. (M-H): 5860.5, Found (M-H): 5860.7

Dibenzocyclooctyne-Modified Oligonucleotide 3.34

To a vial containing 400 µM SMN antisense oligonucleotide containing 3'-amino modifier in nuclease-free H_2O (100 µL, 0.04 µmol, 1 eq.) was added 10 mM

dibenzocyclooctyne-NHS ester in DMF (50 µL, 0.5 µmol, 13 eq.). To this solution was added 20 µL DIPEA, then diluted to 250 µL with DMF. Solution was heated in a thermal cycler to 94°C for 30 minutes, then 72°C for 30 minutes. Analysis by MALDI-MS indicated that desired product was synthesized. LR-MS: calc. (M+H): 4231.9. Found (M+H): 4231.1.



Dibenzocyclooctyne Click Reaction 3.35

To a vial containing DBCO-antisense oligonucleotide in DMF (200 µL, 0.04 µmol, 1 eq.) was added peptide dendron Gly_3-DR-acetyl containing azide linker (3.8 mg, 2.0 µmol, 50 eq.) in 100 µL DMF. Solution was heated in a thermal cycler to 94°C for 30 minutes, then 72°C for 30 minutes. Analysis by MALDI-MS indicated that desired product was synthesized. LR-MS: calc. (M+H): 6126.9. Found (M+H): 6108.5.

3.5 References

(1) Haynes, C.; Iakoucheva, L. M. *Nucleic Acids Res.* **2006**, *34*, 305.

(2) Crespo, L.; Sanclimens, G.; Pons, M.; Giralt, E.; Royo, M.; Albericio, F. *Chem. Rev.* **2005**, *105*, 1663.

(3) Grayson, S. M.; Frechet, J. M. J. *Chem. Rev.* **2001**, *101*, 3819.

(4) Lee, C. C.; MacKay, J. A.; Frechet, J. M.; Szoka, F. C. *Nat. Biotechnol.* **2005**, *23*, 1517.

(5) Fields, C. G.; Mickelson, D. J.; Drake, S. L.; McCarthy, J. B.; Fields, G. B. *J. Biol. Chem.* **1993**, *268*, 14153.

(6) Darbre, T. *Acc. Chem. Res.* **2006**, *39*, 925.

(7) Wu, P.; Feldman, A. K.; Nugent, A. K.; Hawker, C. J.; Scheel, A.; Voit, B.; Pyun, J.; Frechet, J. M.; Sharpless, K. B.; Fokin, V. V. *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 3928.

(8) Thompson, A. S.; Humphrey, G. R.; DeMarco, A. M.; Mathre, D. J.; Grabowski, E. J. J. *J. Org. Chem.* **1993**, *58*, 5886.

(9) Brouwer, A. J., Mulders, S. J. E., Liskamp, R.M.J. *Eur. J. Org. Chem.* **2001**, *2001*, 1903.

(10) Dijkgraaf, I.; Rijnders, A. Y.; Soede, A.; Dechesne, A. C.; van Esse, G. W.; Brouwer, A. J.; Corstens, F. H.; Boerman, O. C.; Rijkers, D. T.; Liskamp, R. M. *Org. Biomol. Chem.* **2007**, *5*, 935.

(11) Arnusch, C. J.; Branderhorst, H.; de Kruijff, B.; Liskamp, R. M.; Breukink, E.; Pieters, R. J. *Biochemistry* **2007**, *46*, 13437.

(12) Branderhorst, H. M.; Kooij, R.; Salminen, A.; Jongeneel, L. H.; Arnusch, C. J.; Liskamp, R. M.; Finne, J.; Pieters, R. J. *Org. Biomol. Chem.* **2008**, *6*, 1425.

(13) Branderhorst, H. M.; Liskamp, R. M.; Visser, G. M.; Pieters, R. J. *Chem. Comm.* **2007**, 5043.

(14) Halkes, S. B.; Vrasidas, I.; Rooijer, G. R.; van den Berg, A. J.; Liskamp, R. M.; Pieters, R. J. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1567.

(15) Graveley, B. R.; Hertel, K. J.; Maniatis, T. *EMBO J.* **1998**, *17*, 6747.

(16) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2004**, *126*, 15046.

(17) Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. *Proc. Nat. Acad. Sci. U. S. A.* **2007**, *104*, 16793.

(18) Jewett, J. C.; Bertozzi, C. R. *Chem. Soc. Rev.* **2010**, *39*, 1272.

(19) Campbell-Verduyn, L. S.; Mirfeizi, L.; Schoonen, A. K.; Dierckx, R. A.; Elsinga, P. H.; Feringa, B. L. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 11117.

(20) Marks, I. S.; Kang, J. S.; Jones, B. T.; Landmark, K. J.; Cleland, A. J.; Taton, T. A. *Bioconjug. Chem.* **2011**, *22*, 1259.

(21) Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. *Anal. Biochem.* **1970**, *34*, 595.

CHAPTER 4

ANALYSIS OF SYNTHETIC SR PROTEINS BY *IN VITRO* SPLICING ASSAY

4.1 Introduction

4.1.1 Analysis of pre-mRNA Splicing

Analysis of the alternative splicing reaction is an important part of determining the effects of splicing modulators. By determining the differences associated with certain molecules, it is possible to identify effective splicing modulators that could lead to potential therapeutic treatments. The products of the splicing reaction, including intermediates and fully spliced final products, can be analyzed by a number of methods that will isolate and visualize the mRNA. The classic method for analysis of mRNA is Northern blotting, in which the mRNA is separated by gel electrophoresis and then transferred to a second membrane where it can interact with antibodies to produce a signal that can be read, either via fluorescence or radioactivity.[1] Newer and more efficient methods of mRNA detection, including quantitative RT-PCR, have made it easier to analyze mRNA splicing products.[2-5] The effects of splicing modulators could initially only be determined by isolating the mRNA from cells or harvested tissues, time-consuming tasks that first require delivery of the molecules to the cells and then isolation of the mRNA. This method creates limitations to the possible molecules that can be analyzed – splicing modulators need to cross the cellular membrane (using a transfection reagent or passive cellular uptake), and then cells or tissues need to be destroyed to harvest the mRNA. Thus, other methods for analysis of splicing modulators would be helpful.

The study of mRNA alternative splicing yielded an assay that could be utilized for analysis of splicing products and the effects of splicing modulators without needing to isolate the mRNA from cells. In the 1980's, Krainer and co-workers began systematically analyzing the products of alternative splicing in order to determine what factors are necessary for the splicing reaction to occur.[6-10] In doing so, they determined what factors could be used to reconstitute the splicing reaction outside of the cell, developing an *in vitro* splicing assay that contained the most important components – ATP and creatine phosphate as the energetic phosphate donor; magnesium ions to help facilitate the proteins that assisted in splicing; a buffer to keep the pH at a consistent value; a size-exclusion polymer such as polyvinylalcohol to increase the effective

concentration of the proteins; and nuclear extracts containing the correct splicing proteins, often isolated from HeLa cells.[11] By adding these components in the proper ratios, one can reconstitute the environment necessary for pre-mRNA splicing to occur *in vitro*, without the need for the cellular conditions.

There are several advantages to using the *in vitro* splicing assay over cellular assays. First and most importantly, splicing modulators can be used without the need for transfection across the cellular membrane. While small molecules may be able to diffuse across the cell membrane, larger molecules with properties that are unfavorable for cellular conditions are more difficult to analyze. These molecules can be analyzed more easily using an *in vitro* assay. The mRNA and proteins necessary for splicing to occur are readily available *in vitro* without needing to cross a cellular membrane. Second, visualization of the mRNA can be much more easily accomplished than with cellular assays, as the mRNA can be transcribed from DNA using a radiolabeled NTP, incorporating a radiolabel that can be visualized after gel electrophoresis by autoradiography or phosphorimaging. This reduces the number of required steps for analysis, compared to cellular mRNA, which must be isolated, converted to cDNA, then amplified by RT-PCR and visualized by some other method. Third, the *in vitro* splicing assay can be easily modified to accommodate a variety of mRNA transcripts.

4.1.2 Development of In Vitro Splicing Assay

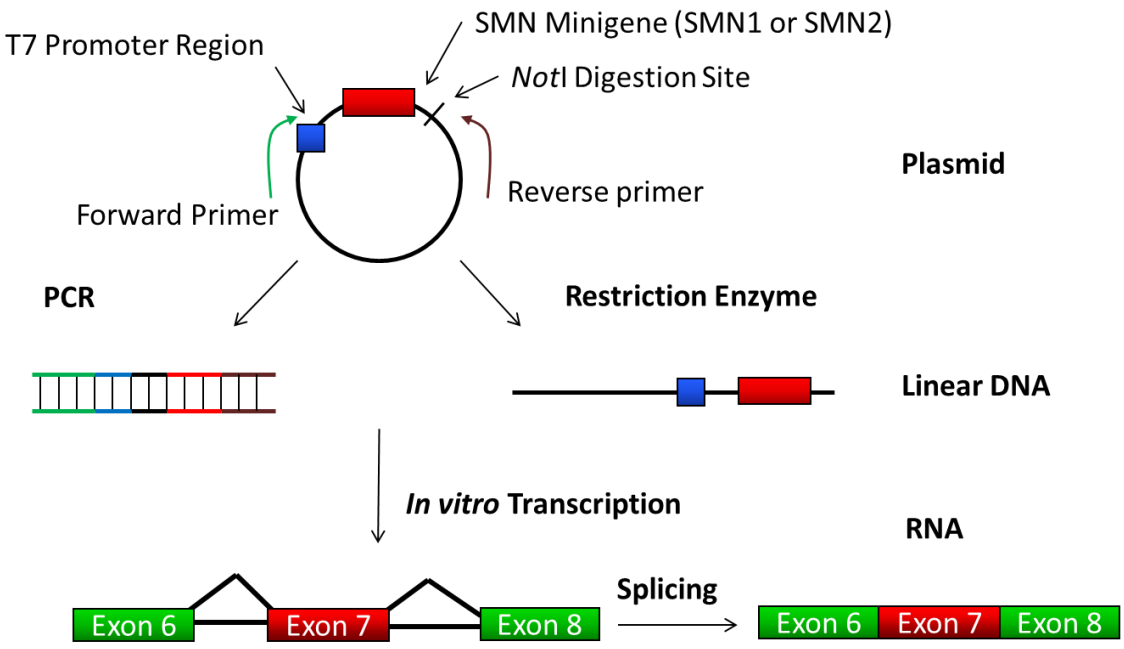


Figure 4.1. *In Vitro* Splicing Assay Project Design

The *in vitro* splicing assay is the final step in a series of reactions that produces mRNA transcripts for the splicing reaction (**Figure 4.1**). First, a DNA plasmid containing the gene of interest must be amplified, either by a miniprep procedure to produce more plasmid or by utilizing polymerase chain reaction (PCR) to amplify the minigene. The plasmid must be further processed by restriction enzyme digestion to linearize the DNA and gel purified to ensure that any remaining proteins or enzymes are removed prior to later applications. Following purification, an *in vitro* transcription reaction incorporating a ^{32}P radiolabeled rUTP produces the pre-mRNA transcript that is universally labeled with a radioactive marker, so that all mRNA products can be visualized by phosphorimaging or autoradiography. Following isolation of the pre-mRNA transcript, the product can be spliced in the presence of our molecules and separated by polyacrylamide gel to visualize the products.

4.2 Results and Discussion

4.2.1 Miniprep

The DNA plasmids containing the *SMN1* and *SMN2* minigenes were graciously donated by Adrian Krainer of Cold Spring Harbor Laboratory. The plasmids contained modified *SMN* gene sequences in which the introns had been significantly shortened to ease *in vitro* synthesis of pre-mRNA.[12,13] In addition, following the first 75 nucleotides of exon 8 a consensus 5' sequence (GTAAGTACTT) was added to improve splicing efficiency of the minigene. The minigene insertion site was surrounded by restriction enzyme digestion sites as part of the ligation insertion procedure, which could be utilized to linearize the plasmid prior to *in vitro* transcription. The miniprep procedure was utilized to increase DNA plasmid concentrations for further manipulations. In this procedure, the plasmid is transformed into an E. coli cell (XL-1B) that is able to produce more plasmid copies. The cells are incubated on LB media supplemented with ampicillin, and only cells containing the desired plasmid and expressing the ampicillin resistance gene are able to survive. Colonies are picked and further amplified by incubation in LB broth. Isolation of the DNA gives a high yield of the plasmid for further manipulations. Plasmids were treated with Proteinase K and gel purified prior to further manipulation to remove any potential RNase contamination and eliminate any side products (**Figure 4.2**).

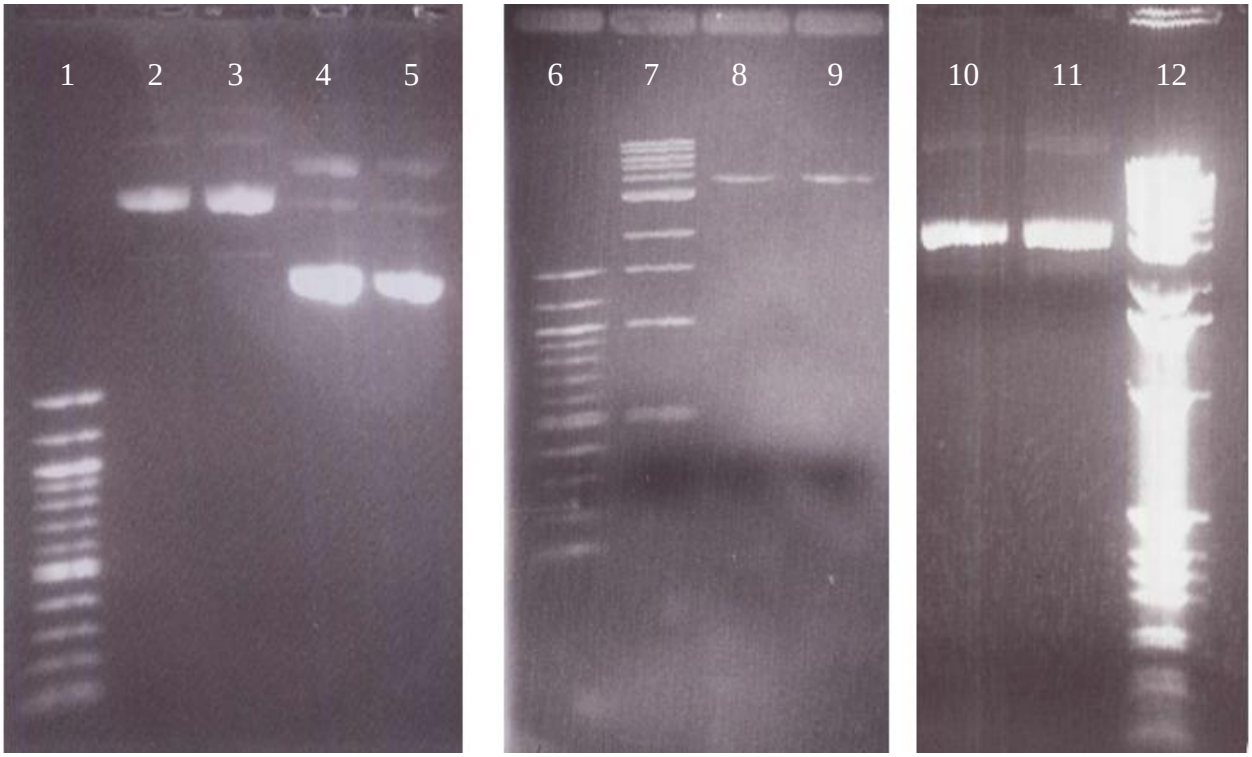


Figure 4.2. DNA plasmids isolated during project. 1) 100 bp ladder (NEB), 2) *SMN1* (Krainer), 3) *SMN2* (Krainer), 4) *SMN6-7-8* (IDT), 5) *SMN6-8* (IDT), 6) 100 bp ladder (NEB), 7) 1 kb ladder (NEB), 8) *SMN1* (Genscript), 9) *SMN2* (Genscript), 10) *Bcl-x* (Caputi), 11) *Bcl-x* (Caputi), 12) 1 kb ladder (NEB)

Sequencing of the plasmids using pCI Forward and pCI Reverse primers targeted to the minigene region of the plasmid revealed some discrepancies between the published sequence and that of our plasmids. In particular, the 5' splice sequence was not included in the plasmids that we were sent. Instead, the plasmids continued on with exon 8 for an extended number of nucleotides, following with the *Not*I enzyme digestion site. Our initial attempts to utilize this plasmid for transcription after linearizing with the *Not*I enzyme were unsuccessful. Thus, we sought several alternatives to solve issues associated with sequence and lack of the correct sequence.

In the latter stages of this project, we were able to obtain synthetic minigenes from Genscript. This company offers customized plasmids containing minigenes of interest for molecular biology manipulations. These plasmids were custom designed to contain the corrected minigene sequence as described in the initial work by Krainer and co-workers,[13] including the consensus 5' splice site sequence. Work with these plasmids was not without its issues. For

example, the plasmid vector used for the custom minigenes did not contain a polymerase promoter, so it was necessary to include the promoter in the custom minigene. The T7 promoter sequence utilized was based off of the sequence contained in the pCI vector (available from Promega), on which our first plasmids were based. Initial work with this plasmid was problematic, as the pCI vector contains a proprietary nucleotide change in the T7 sequence that alters the 3' end of the T7 promoter. Conversations with Promega technical service indicated that this change should not affect transcription of the minigene, but did suggest that a normal T7 promoter primer for sequencing could not be used, clarifying why we were having difficulties obtaining the plasmid sequences.[14] Therefore, special primers targeted to the pUC57 plasmid (the plasmid in which the minigene is synthesized) were required. The sequences that were obtained matched those included in the design. These plasmids were linearized using the *Not*I enzyme as described above (**Figure 4.2**).

Several additional plasmids were prepared by miniprep to be utilized in different ways during the course of our molecule analyses. One issue associated with transcription and splicing of the *SMN1* and *SMN2* mRNA was determining the exact location of the desired splicing products on our gels. Labeling of molecular weight markers was one option, but an additional option was the synthesis of plasmids with minigenes containing the desired correctly and incorrectly spliced sequences, in absence of intervening introns. Transcription of these plasmids provided additional molecular weight markers corresponding to our desired products, making it easier for analysis. In addition, these plasmids could be important for alternate methods of splicing detection.[15] The *SMN6-7-8* and *SMN6-8* plasmids were purchased from IDT. These plasmids incorporated a T7 polymerase promoter sequence and *Not*I enzyme digestion site, making it possible to linearize and transcribe the plasmid as described (**Figure 4.2**).

To examine the effects of our molecules on alternative splicing in a different system, we also used a *Bcl-x* minigene system developed by Caputi and co-workers. This minigene was used for the analysis of synthetic SR proteins for modulating its splicing.[16] The *Bcl-x* gene is spliced into long (Bcl-xL) and short (Bcl-xS) forms, with the different forms playing an important role in cancer cell proliferation. We began using these plasmids in the hopes of showing the effectiveness of our SR protein mimics in a different system, highlighting that these molecules are modular and can work in a variety of contexts. Efforts have been made to switch

the splicing from the anti-apoptotic Bcl-xL form to the pro-apoptotic Bcl-xS form. We isolated the plasmid as described above and linearized it using *Not*I for *in vitro* transcription and splicing (**Figure 4.2**).

4.2.2 PCR

To further amplify the minigene for transcription and to incorporate the desired 5'-consensus sequence into the DNA template, we utilized the polymerase chain reaction (PCR). Our initial attempts to utilize PCR focused primarily on amplifying the minigene site. We used two primers, pCI Fwd and pCR Rev, which were targeted to regions outside of the minigene and incorporated the minigene, enzyme digestion sites, and T7 promoter site (see Materials and Methods). We chose to use the *Pfu* Turbo enzyme rather than another enzyme such as *Taq* polymerase because it was a highly efficient, high fidelity enzyme that was reported to have few errors during replication. A screen of different conditions, primarily focusing on magnesium ion concentration, provided a set of conditions from which a high amount of DNA could be synthesized.

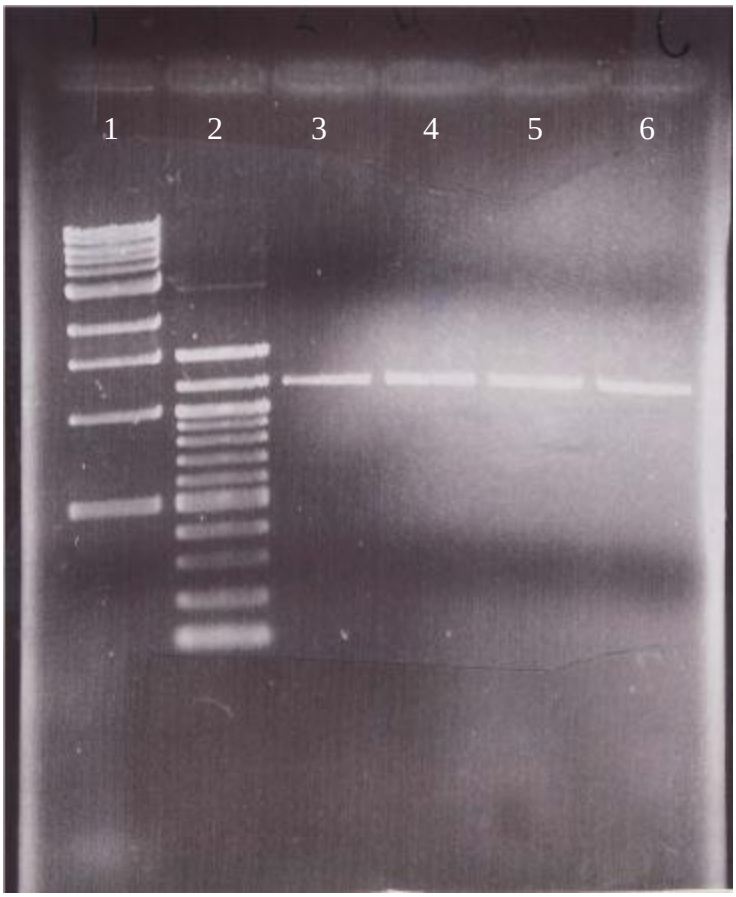


Figure 4.3. PCR products isolated using OneTaq 1) 1 kb ladder (NEB), 2) 100 bp ladder, 3) *SMN1*, 4) *SMN1*, 5) *SMN2*, 6) *SMN2*

As described above, one issue with the plasmid was the lack of a 5' splice site sequence at the 3'-end of the minigene. We determined that we could utilize PCR to incorporate this sequence by directly adding it via a primer. This would not only produce PCR product for transcription, but would provide a substrate for ligation into a new plasmid, should that route be followed. Two primers were developed for PCR, a forward primer that contained the same target as the pCI Fwd but was longer to match the length of the reverse primer, and a reverse primer targeted just downstream of the first 75 nucleotides of exon 7 containing the 5' splice site sequence and two enzyme digestion sites, *Not*I and *Sal*I (see Materials and Methods). Amplification using these primers and the *Pfu* Turbo enzyme required additional modifications, most notably the use of 5% v/v DMSO in the reaction mixture to prevent secondary structure of the primers due to large size and high GC content. Switching the enzyme to several different master mixes, including the EconoTaq Green 2X (Invitrogen) and, more recently, OneTaq (New England Biolabs), provided high yields of the desired PCR products (**Figure 4.3**).

4.2.3 In Vitro Transcription

In vitro transcription is a technique that can be utilized to produce large quantities of long pre-mRNA transcripts for analysis by the *in vitro* splicing assay. In this reaction, a bacterial RNA polymerase enzyme is activated by an enzyme promoter region on the DNA in the presence of buffer and the necessary rNTPs as starting materials for the polymerase reaction. An [α-^{32}P] radiolabeled rUTP can be included in the reaction mixture for universal labeling of the transcript, which can then be visualized by phosphorimaging or autoradiography. In addition, the incorporation of a Ribo m7G cap analog – similar to the cap attached to the 5'-end of endogenous RNA transcripts – can help to improve pre-mRNA stability and splicing, as well as translation should it be further studied.

The development of our *in vitro* transcription reaction required optimization of a number of factors, including reaction conditions and reagents utilized in the reaction. All plasmids contained a T7 promoter region for transcription. The initial reaction conditions were based on the protocol provided by Promega with their T7 RNA polymerase, which was focused on the synthesis of short, high activity probes. Attempts to use this protocol were unsuccessful, and analysis of the protocol helped us to identify several points for optimization. First, reaction time was modified, as the long (~1000 base) pre-mRNA transcript was much longer than the short probes described in the protocol. Analysis of the reaction time from 2 to 8 hours indicated that a reaction time of longer than 4 hours was necessary for sufficient transcript to be produced. Next, the concentrations of the rNTP starting materials were modified so that more nucleotides were available for transcription. 10 mM stocks were used for all rNTPs except rUTP, which was kept at a lower concentration (500 µM) so that radiolabeled rUTP nucleotide could be incorporated into the transcript. Addition of a cap analog as described above further increased the yield and stability of the transcript (**Figure 4.4**).

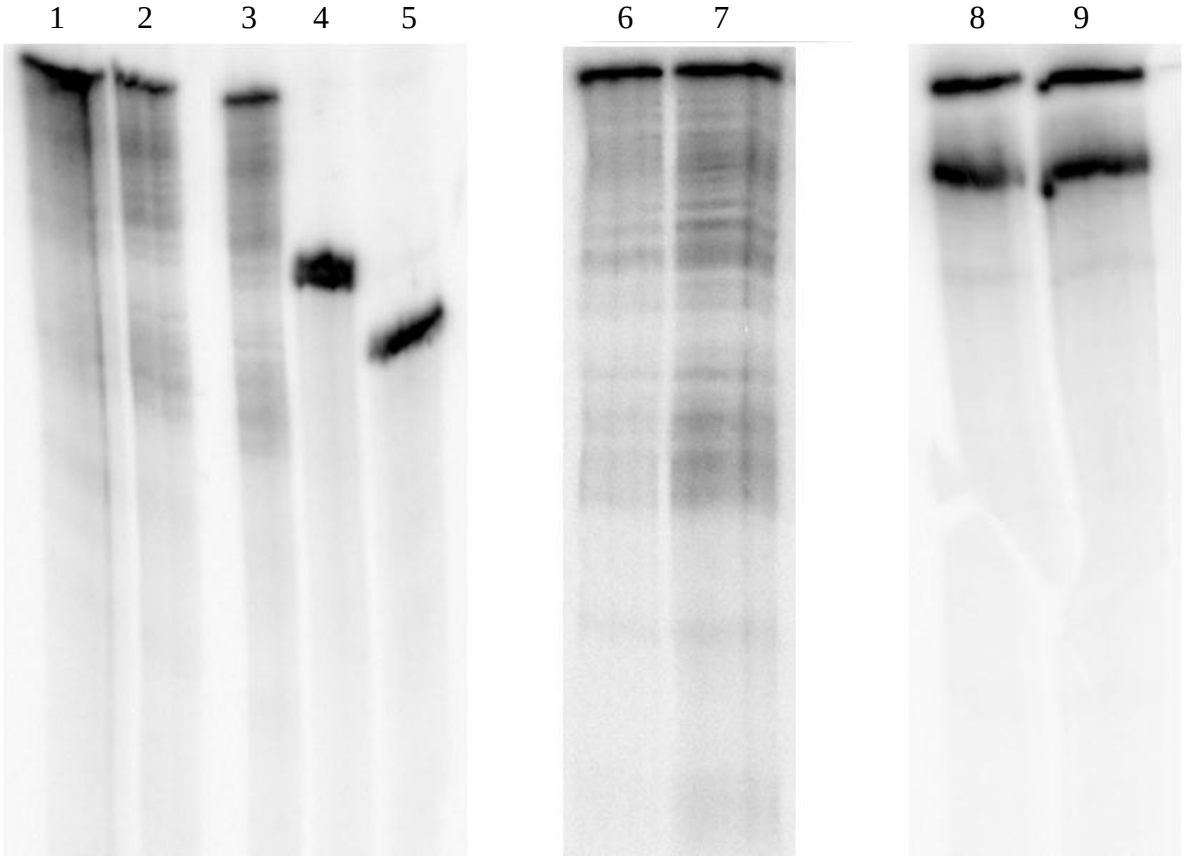


Figure 4.4. *In Vitro* Transcription of DNA templates. 1) *SMN1* (Krainer), 2) *SMN2* (Krainer), 3) *SMN2* (Krainer), 4) *SMN6-7-8* (IDT), 5) *SMN6-8* (IDT), 6) *SMN1* (Genscript), 7) *SMN2* (Genscript), 8) *Bcl-x* (Caputi), 9) *Bcl-x* (Caputi)

Despite optimization of the reaction yielding better results, further work was necessary to increase the amount of desired transcript and decrease side products. A consistent feature of the *in vitro* transcription reaction was the presence of faint bands whose length appeared to correspond to aborted transcription reactions or degradation of full length transcript. Analysis of all reagents for RNase activity indicated that this was not the issue. Purification of the transcripts would be necessary.

Several methods were utilized to assist in transcript purification. The first used Amicon Ultra-0.5 mL Centrifugal filters (Millipore) containing 100 kDa molecular weight cutoff membranes. This strategy would purify the transcripts by allowing smaller side products (smaller than 100 kDa) to pass through the membrane, with the remaining solution containing all products greater than 100 kDa. The estimated number of bases that correspond to 100 kDa is approximately 300 bases, meaning that any products corresponding to the fully spliced minigene

(both exon 7 inclusion and exclusion products are less than 300 bases) would no longer be in the transcript sample. Purification by this method was moderately successful, as smaller products were indeed removed from the mixture as indicated by gel electrophoresis, and the yield from the purification was very good. Not all side products were removed, however, and further analysis during splicing was still somewhat marred by the presence of side products. A more thorough method for purification was still needed.

A standard method for purification of transcripts is to use gel purification to remove and isolate the desired bands.[17,18] There are a number of methods available for isolation of RNA from polyacrylamide gel, which use buffers ranging from pure water to sodium acetate buffers containing SDS, and incubate samples at 4°C, room temperature, 37°C, or even higher temperatures. A number of methods were used to purify the *SMN* transcripts, which were only moderately successful and generally only produced a small amount of product that could not be used in splicing assays. The most successful protocol utilized a freeze-thaw method to rapidly crystallize the gel and break it apart to elute the RNA.[18] In this method, the gel pieces are suspended in the elution buffer and rapidly frozen to -80°C using a dry ice bath. The solution is then rapidly heated to 90°C in order to heat shock the gel piece, allowing the transcript to be eluted into the elution buffer. Filtration through a 0.2 µm filter is used to remove the gel pieces remaining in solution, and the RNA is then precipitated from the buffer. As mentioned, this method was moderately successful, and produced the most transcript of all gel purification methods. Unfortunately, use of the transcript in subsequent *in vitro* splicing assays was not successful, and subsequent *in vitro* splicing assays used spin column-purified transcript.

It is interesting to note that the transcription of the *SMN6-7-8*, *SMN6-8*, and *Bcl-x* plasmids appeared to occur with few of the problems associated with *SMN1* and *SMN2* (**Figure 4.4**). Analysis of the plasmids by polyacrylamide gel generally produced a single desired band for each of these plasmids, without the same patterns of shortened bands associated with the *SMN1* and *SMN2* transcripts. It is possible that the complexity and length of the *SMN1* and *SMN2* minigenes may have caused inefficient transcription. This is in contrast with the *SMN6-7-8* and *SMN6-8* plasmids, which both contained no introns. The relatively simple splicing pattern and shorter sequence of the *Bcl-x* plasmid may have made it more stable to degradation. We also speculated that the donated plasmids may have been of lower quality than commercially

available counterparts. However, *SMN1* and *SMN2* plasmids derived from the pUC57 vector and obtained from a commercial source had similar problems to the original *SMN* plasmids. The sequence used for these newer minigenes matched the genomic sequence, making it unlikely that some inherent issue associated with the donated *SMN1* and *SMN2* plasmids was transferred to the new plasmids. Therefore, it is possible that the sequences of *SMN1* and *SMN2* themselves may have played some detrimental role in transcription analysis.

4.2.4 In Vitro Splicing Assay

The use of the *in vitro* splicing assay to analyze the effects of different molecules is well documented, and we used this method to analyze our synthetic SR protein mimics.[11,19-21] The initial design of the assay was based on the work of Mayeda and Krainer, who first developed a standard splicing assay protocol.[11] This general assay reconstitutes the conditions of the nucleus of HeLa cells in order to splice the pre-mRNA. Our initial attempts to use the splicing assay for analysis of *SMN1* and *SMN2* splicing using this method were moderately successful, producing some changes in the arrangement of bands on the gel but few discernible differences between the mRNA splicing products. We initially examined the splicing of unlabeled *SMN1* and *SMN2* pre-mRNA transcripts, with the hope that we would be able to easily identify the desired products without need for a radiolabel. Unfortunately, tRNA in the splicing stop solution (see Materials and Methods) appeared to block the region of interest (200-400 nucleotides) on the agarose gels. Use of a Proteinase K stop solution in substitution for the splicing stop solution did not resolve the issue, so it was decided that a radiolabel must be incorporated into the transcripts in order to ensure that we could more easily visualize the products. Subsequent assays utilized the [α-^{32}P] radiolabeled rUTP nucleotide during transcription, to ensure that the gel could be visualized by phosphorimager analysis after splicing occurred.

A number of steps were taken to optimize the assay prior to the analysis of our synthetic SR protein mimics. The concentrations of several key components in the splicing assay were altered to better match the conditions necessary for splicing of *SMN* minigenes. For example, the $MgCl_2$ concentration in the standard protocol was 3.2 mM, whereas for *SMN* genes it was 1.6 mM according to later protocols.[21] Preparation of fresh reagents was also important, as the activity of certain reagents (especially $MgCl_2$ and ATP/creatine phosphate mix) decreased over time, inhibiting the ability of the nuclear extract proteins to fully activate splicing. The HeLa

nuclear extracts themselves may have been one of the primary reasons for lack of activity. Even when prepared fresh in the lab, these extracts do not display consistent splicing activity, indicating that multiple reactions must be run under the same conditions in order to determine level of splicing activity and obtain consistent results. A survey of different commercially available extracts was conducted to determine which would provide the most consistent activity, and those from Promega appeared to be the most consistent. Previously, Promega had actually produced and sold splicing-grade nuclear extracts, but unfortunately these extracts are no longer available.

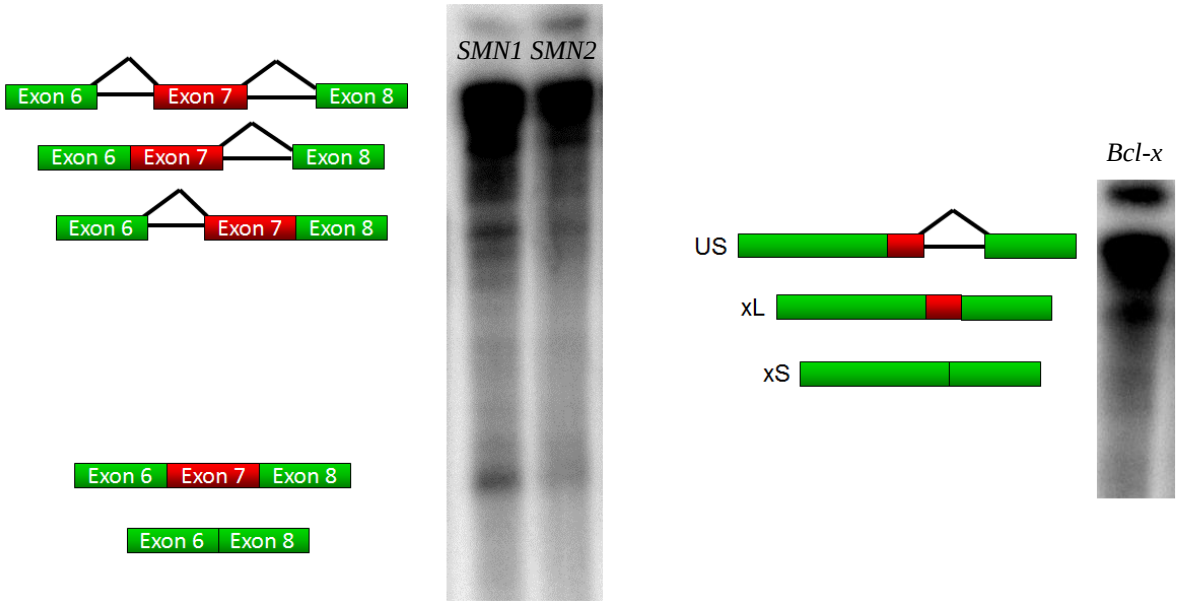


Figure 4.5. *In Vitro* Splicing Assay in the absence of molecules. Possible arrangement of splicing products, as suggested by molecular weight markers, is indicated next to gel. US = unspliced.

Analysis of radiolabeled *SMN1*, *SMN2*, and *Bcl-x* transcripts was conducted in the absence of the peptide-oligonucleotide conjugates to determine how the different transcripts would be spliced under our assay conditions (**Figure 4.5**). Qualitative analysis of the *SMN1* and *SMN2* plasmids showed that some of the desired bands were associated with splicing patterns in previously published work, as indicated by proximity to radiolabeled DNA markers. In spite of our efforts, we were unable to identify bands corresponding to the desired splicing pattern differences between *SMN1* and *SMN2* mRNA. It appeared as though the splicing reaction did not fully complete, producing only bands corresponding to the intermediates. On the other hand, analysis of the *Bcl-x* transcripts appeared to be more successful, producing bands that

corresponded to the Bcl-xL isoform of the mRNA. Very little of the Bcl-xS product was synthesized, as expected. Although the *Bcl-x* plasmid appeared to be much better for analysis, we did not perform many additional assays with it.

Despite difficulties with the production of consistent splicing assay results, we continued forward and began analyzing the effects of different molecules on *SMN2* pre-mRNA splicing. We hoped to determine the effects of several different molecules on alternative splicing – first, the separate antisense oligonucleotide and peptide or dendrimer components, and then the full conjugates containing both pieces. We hypothesized that the peptide or Dendron components separately would have no effect on splicing, as they should not bind directly to the mRNA. The antisense oligonucleotide was likely to show some activity, as it was shown previously by Krainer and co-workers to have an effect on splicing in several studies.[19,21] We had hoped, however, that the addition of the synthetic RS domain would have an additional beneficial effect on alternative splicing of *SMN2*, producing the most full-length *SMN2* mRNA following the splicing reaction. Analysis of the peptide-oligonucleotide conjugate containing 5 DR dipeptides, along with the separate components of the conjugate, revealed that the conjugate had the most dramatic effect on alternative splicing, producing primarily an mRNA band corresponding to the *SMN6-7-8* molecular weight marker (**Figure 4.6**). Interestingly, there did not appear to be a

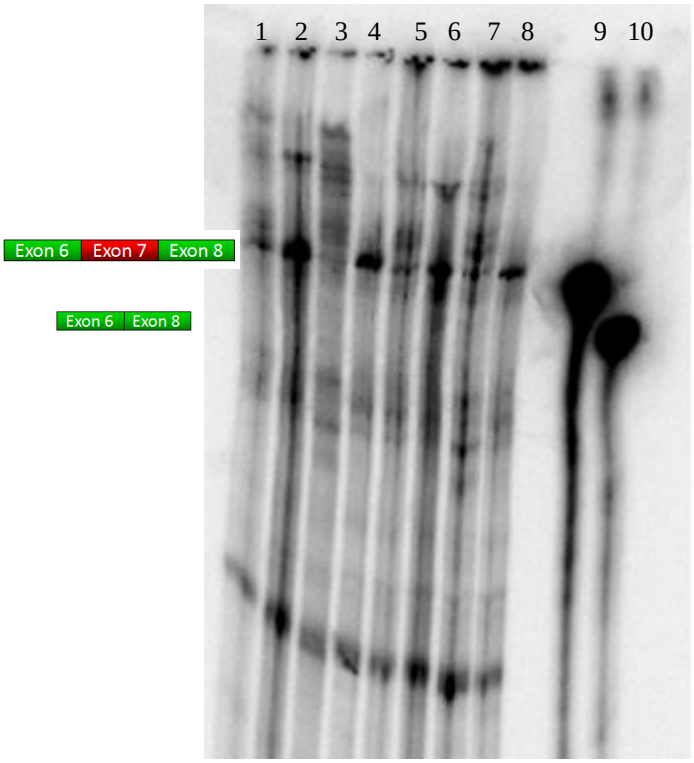


Figure 4.6. Analysis of peptide-oligonucleotide conjugate components on splicing of *SMN1* (lanes 1-4) and *SMN2* mRNA (lanes 5-8). Bands corresponding to correctly and incorrectly spliced *SMN2* are marked. 1) H_2O, 2) *SMN* antisense oligonucleotide, 3) $(DR)_5$ peptide, 4) $(DR)_5$ peptide-oligonucleotide conjugate, 5) H_2O, 6) *SMN* antisense oligonucleotide, 7) $(DR)_5$ peptide, 8) $(DR)_5$ peptide-oligonucleotide conjugate, 9) *SMN6-7-8*, 10) *SMN6-8*

band corresponding to the *SMN6-8* molecular weight marker in the lanes containing the antisense oligonucleotide or peptide-oligonucleotide conjugate.

The results from this first analysis highlighted one important issue that we spent a great deal of time attempting to resolve – the lack of a band corresponding to incorrectly spliced *SMN2*. Analysis of a number of gels showed essentially no band corresponding to the proposed *SMN2* band location, as indicated by both *SMN6-8* transcript and a radiolabeled RNA marker (**Figure 4.6**). A common feature of our gel analysis was the presence of two bands, one corresponding to the *SMN6-7-8* band and one not far below it, still above the *SMN6-8* band, indicating that some reaction had occurred but not any that corresponded to the molecular weight markers. Sequencing of the *SMN6-7-8* and *SMN6-8* plasmids indicated that our proposed number of nucleotides should have corresponded correctly, and they appeared to match well with bands located in the radiolabeled RNA marker. As a further determination of our splicing products, we ran unlabeled splicing assays using *SMN2* and submitted the final products for MALDI-MS. Analysis of the splicing products produced a peak that corresponded with the calculated incorrectly spliced *SMN2* molecular weight, suggesting that this product was forming, although the amount of product being formed was not discerned by this method. Therefore, for subsequent analysis of the splicing assays, the two bands corresponding to *SMN6-7-8* and the band below it were used to determine if the molecules had activity. Although imperfect, this method served as a means of determining molecule effects while alternative methods of analysis were explored.

We next examined the peptide-oligonucleotide conjugates at decreased concentrations to determine the dose response of the molecules. In previous studies, both antisense oligonucleotides and peptide-oligonucleotide conjugates were analyzed at nanomolar concentration, so we examined our molecules at the same concentration in order to be able to compare their effects more directly to previously described molecules.[19,21,22] The concentration of the molecules that were tested was measured using UV spectroscopy, based on the UV signal of the oligonucleotide. Samples were then diluted to a series of concentrations ranging from 400 nM to 1 nM. We hypothesized that decreasing the concentration would decrease the ability of the molecules to modulate splicing.

Qualitative analysis of the molecules revealed that several of the molecules were active at 400 nM concentration, with that activity disappearing as the concentration decreased. In each case, a band approximately the same size as the *SMN6-7-8* marker (corresponding to the correctly spliced *SMN2* mRNA) formed at the highest concentration, with the signal for that band fading at lower concentrations (**Figure 4.7**). Attempts to quantify these changes using the equation for determining *SMN* exon 7 inclusion (see Materials and Methods) indicated that the $(DR)_5$ peptide-oligonucleotide conjugate had the best activity, increasing exon 7 inclusion to nearly 80% as opposed to a background level of around 50%. The DR dendron, containing four sets of DR dipeptide repeats, could only activate splicing to around 70% inclusion. This suggests one of two possible reasons for the change. Either the additional DR dipeptide of the peptide-oligonucleotide conjugate helped to further increase activation, or the arrangement of the charged residues on the periphery, while somewhat effective, does not have as large of an effect when displayed in a nonlinear fashion. Interestingly, analysis of the antisense oligonucleotide appeared to have little effect on *SMN2* splicing at lower concentrations, suggesting further that the presence of both the antisense oligonucleotide and the synthetic RS domain is necessary to create a strong splicing activator when targeted to this sequence on *SMN2*.

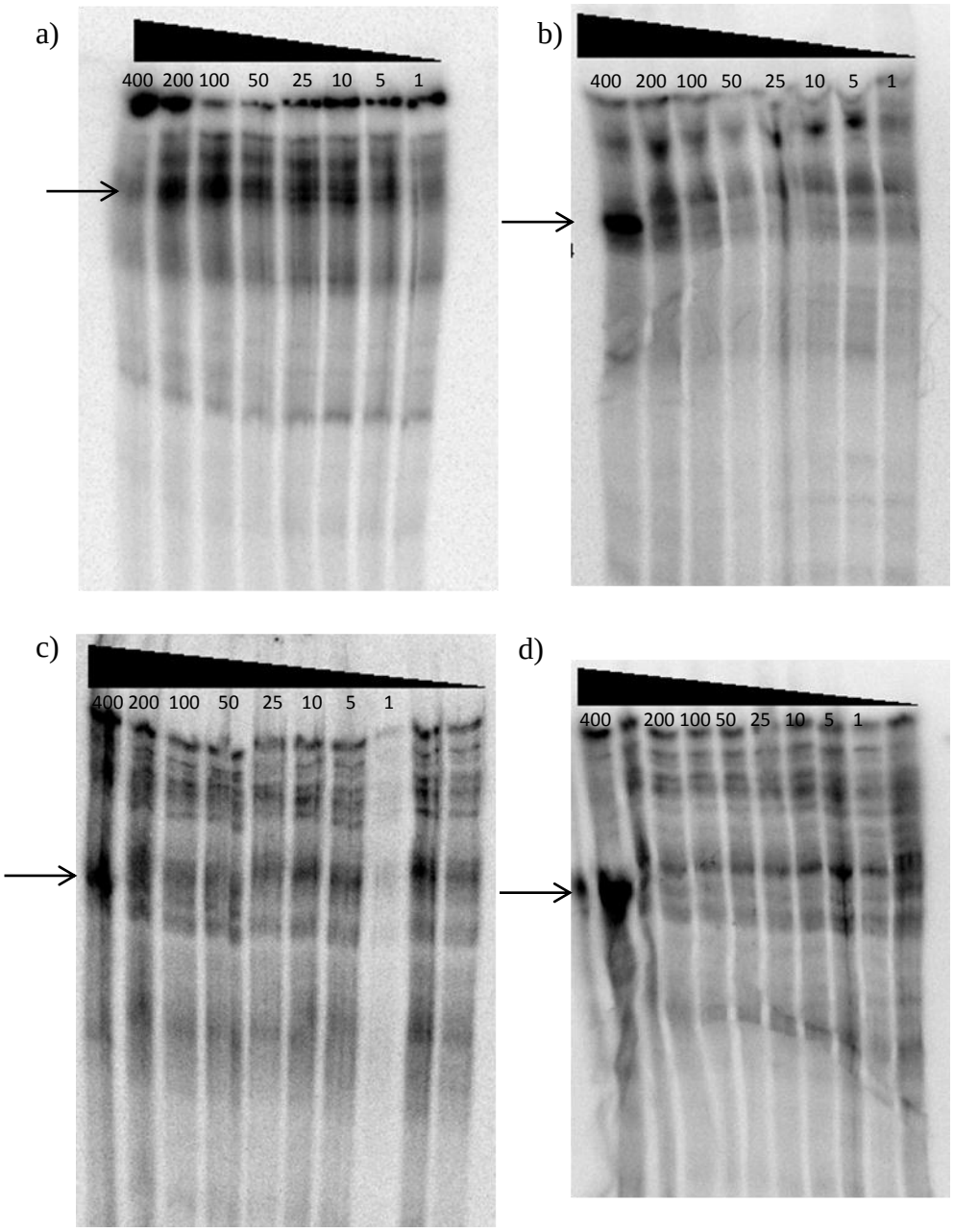


Figure 4.7. Splicing assay gel results. Assays were run as described, using *SMN2* pre-mRNA transcript and serial dilution of 400 nM → 1 nM molecule. Unmarked lanes contain H_2O as a control instead of conjugate molecule. Desired band corresponding to *SMN6-7-8* is highlighted with an arrow. a) Antisense oligonucleotide. b) $(DR)_5$ peptide-oligonucleotide conjugate. c) $(DR)_6$ peptide-oligonucleotide conjugate. d) DR peptide dendrimer-oligonucleotide conjugate.

The major issue associated with analysis of our conjugates was the large amount of inconclusive data that could be obtained from splicing assay gels. In spite of our efforts to optimize both assay and analysis conditions, there was never a clear result from which we could

definitively state that our molecules were having an effect on splicing. Even when using plasmids that had been redesigned to contain the correct sequence elements, we were not able to consistently obtain conclusive results. Use of the *Bcl-x* plasmid was the logical first step in determining what effect the molecules have, as the number of splicing products is significantly smaller and less complex than for *SMN* plasmids, and the unspliced transcript has fewer side products that can disrupt splicing analysis. Lack of time and focus on making the *SMN* plasmids successful prevented further analysis of *Bcl-x*. A second way to improve analysis was to use other alternative splicing analysis methods in order to determine if the desired products are even being formed. It was possible that the analysis by polyacrylamide gel electrophoresis was suboptimal and the biggest cause of our problem, so verification by an alternative method would be helpful.

An alternative detection method for the analysis of splicing assay products uses a set of molecules called 'junction probes'. Developed by Sintim and co-workers in 2008, this technique uses a tripartite array of three molecules – two antisense oligonucleotides targeted to the desired DNA or RNA target and to each other, and the third DNA or RNA target.[15] Upon formation, these molecules bind to the target oligonucleotide and to each other, creating an enzyme digestion sequence that is only available in the presence of the desired target. Enzymatic cleavage of the newly formed sequence releases a quencher on the 5' end of one of the two antisense oligonucleotides, allowing a fluorophore on the other antisense oligonucleotide to fluoresce. Analysis of the fluorescence signal can then be quantified, allowing for the determination of the amount of correct target available in a reaction. These molecules are advantageous in that they are sequence specific and have been developed to contain structure elements to prevent non-specific binding and fluorescence in the absence of a target.[23]

Our work with the Sintim group in development of these molecules for targeting *SMN* splicing products involved synthesizing the *SMN* transcripts while they developed their detection molecules. Initial analysis using the *SMN6-7-8* and *SMN6-8* transcripts showed that they could obtain a strong reading above the background for the *SMN6-7-8* transcript, but not for the *SMN6-8* transcript. Further analysis using DNA templates of the same sequence indicated that they could be targeted to those sequence elements, but not to the RNA. They hypothesized that possible RNA secondary structure may have been interfering with their analysis. It is interesting

to note that they too had issues with identifying the incorrectly spliced *SMN2* mRNA, further suggesting that there may have been issues with the *SMN2* sequence or plasmids. Unfortunately, further work into this matter was not conducted during the latter stages of this project.

4.3 Conclusions and Future Directions

The *in vitro* splicing assay is an important technique that can be used to analyze the effects of splicing modulators under acellular conditions using HeLa cell nuclear extracts. This is advantageous over cellular assays in that the molecules do not need to traverse the cell membrane, and that the products can be analyzed directly on polyacrylamide gels if they incorporate a radiolabeled nucleotide. Plasmids containing *SMN1*, *SMN2*, *SMN6-7-8*, *SMN6-8*, and *Bcl-x* minigenes were amplified by Miniprep and either linearized using restriction enzymes or further amplified using PCR. These plasmids were then transcribed into pre-mRNA using an *in vitro* transcription reaction. Depending on the starting material, the products of the reaction either consisted of a single band or multiple bands that required purification by spin column filtration or elution from the gel. Alternative splicing of these transcripts was analyzed using the *in vitro* splicing assay. Using this assay, qualitative differences were observed in the *SMN1* and *SMN2* mRNA following splicing in the presence of our conjugates. A number of conjugates were tested, but many remain to be tested. Despite numerous attempts at optimization, the assay was not conclusive, although trends did appear that make us hopeful that the molecules have activity.

There are a number of ways in which this project could be improved so that a clearer picture of how these molecules affect splicing could be produced. The most obvious would be to use the *Bcl-x* plasmid as the primary target for splicing activation. Splicing of the *Bcl-x* minigene is simpler, containing only two splicing products, and the sequence of the minigene appears to be amenable to transcription and splicing. Use of the *SMN* minigenes, while clinically relevant and able to provide very interesting data if successful, was riddled with problems even when resynthesized using commercially available sources. Further improvements to the splicing analysis would be the isolation and analysis of HeLa cell nuclear extracts in the laboratory, rather than relying on commercial sources. Extract batches could be tested for splicing activity and discarded if not up to desired standards. The facilities for growing and maintaining HeLa cells are generally available. A third way to improve the outcome of the assay, should the *SMN*

minigenes be pursued, would be to determine the necessary sequences for minigene splicing and then to purchase these minigenes from a commercially available source, such as Genscript. Although we had hoped that following this strategy would be useful to us, we based the sequence on the already established minigene provided by Krainer and co-workers. This may have been problematic and one of the causes of our lack of successful results. Re-synthesis of a new set of plasmids that contained the verified correct sequence may have assisted in improving transcription and splicing assay results.

Additional improvements to the project could be made through the use of alternative techniques to examine splicing. A number of different methods are available, and each could be utilized to verify the effects of our molecules. As mentioned above, junction probes were an experimental method for analysis of DNA and RNA, and more work using this technique could provide a quantitative measure of different splicing products. RT-PCR and qPCR are a second set of techniques that could be used to determine the differences in amount of splicing product. The spliced RNA products would first be converted to cDNA molecules using reverse transcriptase, and primers specific to the different spliced products could be amplified and quantitatively measured using fluorescence to determine the different amounts of each product. Attempts were made near the end of this project to use this technique, but the results were not successful, indicating that much more optimization would be necessary to make it a viable technique. Campus facilities for RT-PCR and qPCR would be able to assist with assay development, although early attempts to plan a possible analysis with them were less than encouraging. Development of the primers for qPCR would be the first important hurdle, following which optimization of the assay would lead to the desired products. A third possible assay is the RNase protection assay.[24] Antisense molecules targeted to specific regions of the splicing products, such as the exon-exon junctions, would first be bound to the RNA. Digestion of the remaining splicing products would be completed using an RNase enzyme, following which visualization of the products would reveal the relative amounts of correctly and incorrectly spliced RNA. Generally, this technique uses a radiolabeled antisense oligonucleotide, but it could also be conducted with the radiolabeled RNA transcripts, which have already been synthesized.

4.4 Materials and Methods

General

All buffers for biological experiments were prepared with autoclaved Millipore or Nuclease-free (BioRad) water, pipette tips, and microfuge tubes. T7 RNA polymerase was purchased from Promega. HeLa nuclear extracts were purchased from Promega, Millipore, AnaSpec, and Calbiochem. *Pfu* Turbo was purchased from Stratagene. EconoTaq Plus GREEN 2X Master Mix was purchased from Lucigen. OneTaq 2X Master Mix was purchased from New England Biolabs. Plasmid DNA concentration was measured using a UV spectrometer. Plasmid DNA was sequenced at the Core Sequencing Facility, Keck Institute, University of Illinois. Antisense oligonucleotides were purchased from Yale Keck Oligonucleotide Synthesis Facility using modifiers purchased from Glen Research. Agarose was purchased from Fischer Scientific. Denaturing acrylamide gel solutions were purchased from National Diagnostics. Acrylamide gels were visualized on phosphorimaging screens (Molecular Dynamics) and were scanned using a Molecular Dynamics Storm 240 Scanner. Gel bands were quantified using ImageQuant software (Molecular Dynamics).

Plasmid Transformation and Miniprep

5 µL of the desired plasmid in buffer EB was mixed with 50 µL competent XL-1B cells and incubated on ice for 30 minutes. Cells were heat shocked for 90 sec at 42°C and immediately put on ice. 200 µL SOC buffer was added, and cells were incubated at 37°C with shaking for 60 min. 35 µL cells were plated onto LB media containing Ampicillin and incubated at 37°C for 16 hours. Single bacterial colonies were transferred to 5 mL LB media containing 5 µL Ampicillin (1000x) and incubated with shaking at 37°C for 16 hours. Plasmid DNA was then isolated using a Qiagen Miniprep Kit (Qiagen, CA). Plasmid DNA samples were gel purified on 1% agarose gels using 0.5X TBE buffer containing 0.5 mg/mL ethidum bromide. Desired bands were removed and isolated using a Qiagen Gel Extraction Kit (Qiagen, CA). Gels were visualized on a BioRad Gel Imager. Isolated plasmid DNA from each sample was stored in 1.5 mL Eppindorf tubes at -20°C.

Plasmid Sequences

Plasmids containing *SMN1* and *SMN2* minigenes were graciously donated by Dr. Adrian Krainer, Cold Spring Harbor Laboratory or purchased from Genscript.[13] Plasmid containing *Bcl-x* minigene was graciously donated by Dr. Massimo Caputi, Florida Atlantic University.[16] Plasmids containing *SMN6-7-8* and *SMN6-8* minigenes were purchased from IDT.

SMN6-7-8

CCCGAGATTCCCCCACCACCTCCCATATGTCCAGATTCTCTTGATGATGCTGATGCTT
TGGGAAGTATGTTAATTTCATGGTACATGAGTGGCTATCATACTGGCTATTATATGG
GTTTTAGACAAAATCAAAAAGAAGGAAGGTGCTCACATTCCTTAAATTAAGGAGAA
ATGCTGGCATAGAGCAGCACTAAATGACACCACTAAAGAAACGATCAGACAGATCT
GGAATGTGAAGCGTTATAGGAATGTGAAGCGTTATAGACGATAACTGGCCT

SMN6-8

CCCGAGATTCCCCCACCACCTCCCATATGTCCAGATTCTCTTGATGATGCTGATGCTT
TGGGAAGTATGTTAATTTCATGGTACATGAGTGGCTATCATACTGGCTATTATATGG
AAATGCTGGCATAGAGCAGCACTAAATGACACCACTAAAGAAACGATCAGACAGAT
CTGGAATGTGAAGCGTTATAGGAATGTGAAGCGTTATAGACGATAACTGGCCT

Restriction Enzyme Digestion

Restriction enzyme digestion was performed using *Not*I and *Xho*I restriction enzymes using a modified version of the manufacturer's protocol (New England Biolabs). Briefly, each reaction contained 10 µL nuclease-free H_2O, 2 µL Buffer 3 (NEB), 1 µL bovine serum albumin, 8 µL DNA template, and 2 µL restriction enzyme. Samples were incubated at 37°C overnight. Following incubation, digested plasmids were separated by 1% agarose gels and purified using Qiagen Qiaquick PCR Purification Kit (Qiagen, CA) or IBI Gel/PCR DNA Fragment Extraction Kit (IBI Scientific, IA). Products were gel purified and stored as described previously.

PCR

PCR was performed using a MJ Research, Inc. PTC-100 Programmable Thermal Controller or a Techne TC-3000 Thermal Cycler using gel purified *SMN* plasmids, according to

manufacturer's protocols (primer sequences, cycling times and temperatures are given below), except 5% (v/v) DMSO was added to each reaction. PCR products were purified using a Qiagen Qiaquick PCR Purification Kit (Qiagen, CA) or IBI Gel/PCR DNA Fragment Extraction Kit (IBI Scientific, IA). Products were gel purified and stored as described previously.

In vitro Transcription

In vitro transcription was performed using T7 RNA polymerase using a modified version of the manufacturer's protocol for high activity probes (Promega). Briefly, each reaction contained 4 µL Transcription-optimized 5x buffer, 2 µL 100 mM DTT, 1 µL RNase inhibitor (Promega), 4 µL 10 mM rNTP-rUTP mix, 2 µL 500 µM rUTP, 1.0 µL 40 mM Ribo m^7G cap analog (Promega), 4 µL DNA template, 2 µL [α-^{32}P]-labeled rUTP, and 2 µL T7 RNA polymerase. Samples were incubated for 2 hr at 37°C. To each reaction was then added 4 µL 10 mM $CaCl_2$ and 2 µL RNase-free DNase (New England Biolabs), and reactions were incubated at 37°C for 30 min. To each reaction was then added 2 µL EDTA, pH 8.0, 6 µL 5M NaCl, and 150 µL TE buffer, pH 8.0. 200 µL 25:24:1 Phenol:$CHCl_3$:isoamyl alcohol was added, and tubes were vortexed briefly and centrifuged at 13,000 rpm for 1 min. The organic layer was removed, and the extraction repeated with 200 µL 24:1 $CHCl_3$:isoamyl alcohol. To the remaining aqueous layer was added 500 µL EtOH, and tubes were stored at -20°C overnight. Samples were then centrifuged at 13,000 rpm for 20 min, the supernatant was removed, and samples were dried using a speed vacuum. Samples were then resuspended in 40 µL TE buffer, pH 8.0 for further use. All RNA samples were denatured prior to analysis on 8% denaturing acrylamide gels by the addition of a formaldehyde-formamide loading buffer followed by heating to 85°C for 2 minutes. *Spin Column Purification*: RNA transcripts were diluted to 500 µL and transferred to an Amicon Ultra 0.5 mL Centrifugal Filter with 100 kDa molecular weight cutoff membrane (EMD Millipore, MA). Samples were centrifuged for 20 minutes. Remaining solvent in upper chamber was collected in new vials and analyzed by gel electrophoresis. *Gel Purification*: Gel pieces containing desired transcripts were removed and suspended in gel extraction buffer. Samples were frozen to -80°C using a dry ice bath, then rapidly heated to 90°C with a hot water bath. Samples were then incubated overnight with shaking at 37°C. Following incubation, samples were passed through 0.2 µm filters to remove gel pieces. Buffer was extracted and EtOH

precipitated as described previously, and transcripts were stored in TE buffer as described previously.

In Vitro Splicing Assay

In vitro splicing assays using *SMN* plasmids were performed using a modified version of a protocol outlined by Krainer and co-workers. Briefly, each reaction contained 0.5 µL 25x ATP/CP mix, 0.26 µL 80 mM $MgCl_2$, 0.64 µL 0.4 M HEPES-KOH, pH 7.3, 0.5 µL RNase inhibitor (Promega), 2.6 µL nuclease-free H_2O, 2.5 µL 13% polyvinyl alcohol (PVA), 4.0 µL HeLa nuclear extract, and 3.0 µL transcribed RNA. *In vitro* splicing assays using *Bcl-x* plasmids were performed using a protocol outlined by Caputi and co-workers. Briefly, each reaction contained 4.8 µL 13% PVA, 1.0 µL 100 mM $MgCl_2$, 1.0 µL 100 mM ATP, 1.0 µL 125 mM CP, 0.5 µL RNase inhibitor, 7.5 µL HeLa nuclear extract, 7.5 µL nuclease-free H_2O, and 3.0 µL transcribed RNA. For assays analyzing molecules, 0.5 µL molecule diluted in nuclease-free H_2O was added to reaction mixture prior to incubation. 0.5 µL nuclease-free H_2O was added for control reactions. Concentrations are given with gel analysis. Samples were incubated for 4 hr at 37°C. At that time, 200 µL Splicing Stop Solution or Proteinase K (see SI) buffer was added to each sample. Samples were then extracted as described above, except Tris-buffer saturated phenol (Invitrogen) was substituted for 25:24:1 Phenol:$CHCl_3$:isoamyl alcohol. Samples were EtOH precipitated and resuspended in 5 µL RNA dye and 5 µL TE buffer, denatured by heating to 85°C for 2 minutes, then analyzed on 8% denaturing acrylamide gels as described above. Exon7 inclusion was calculated as a percentage of the total amount of fully spliced mRNA, i.e. included mRNA X 100/(included mRNA + skipped mRNA).

Buffers

10 X TBE Buffer (1 L):

- 108 g Tris Base

- 55 g Boric Acid

- 9.3 g Disodium EDTA

- Stored at 4°C

Formaldehyde-Formamide RNA Buffer:

- 2 mL Formamide

- 700 µL 37 % Formaldehyde

- 400 µL MOPS buffer

- Distributed in aliquots of 20 µL and stored at -20°C

MOPS Buffer:

- 0.2 M MOPS, pH 7

- 50 mM NaOAc

- 5 mM EDTA, pH 8

RNA dye

- 90% (v/v) formamide

- 50 mM Tris-HCl, pH 7.5

- 1 mM EDTA

- 0.1% bromophenol blue

- 0.1% xylene cyanol FF

- Stored at 4°C

Acrylamide Gel Extraction Buffer

- 10 mM Tris, pH 7.5

- 1 mM EDTA

- 0.5 M Ammonium Acetate, pH 5.6

- 0.1% (w/v) SDS

- Stored at 4°C

Splicing Stop Solution

- 0.3 M Sodium Acetate, pH 5.2

- 0.1% (w/v) SDS

- 62.5 µg/mL tRNA

- Stored at 4°C

Proteinase K Buffer

- 10 mM Tris

- 1% (w/v) SDS

- 0.15 mM NaCl

- 10 mM EDTA

- 0.25 mg/mL glycogen

- 0.25 mg/mL Proteinase K

- Stored at 4°C

Buffer D

- 20 mM HEPES-KOH, pH 8.0

- 100 mM KCl

- 0.2 mM EDTA

- 20% (v/v) glycerol

- 1 mM DTT

- Stored at 4°C

PCR Primers

Forward: 5'-
GCTAACGCAGTCAGTGCTTCTGACACAACAGTCTCGAACTTAAGCTGCAGAA-3'

Reverse: 5'-GAAGCGGCCGCGTCGACAAGTACTTACCTGTAACGCTCCACATTCCAG-3'

PCR Cycling Times and Temperatures

94°C- 2 minutes

94°C- 30 s (denaturation)

55°C- 30 s (annealing)

72°C- 2 min (extension)

30 cycles

72°C- 10 min (finishing)

4.5 References

(1) Kevil, C. G.; Walsh, L.; Laroux, F. S.; Kalogeris, T.; Grisham, M. B.; Alexander, J. S. *Biochem. Biophys. Res. Commun.* **1997**, *238*, 277.

(2) Coady, T. H.; Baughan, T. D.; Shababi, M.; Passini, M. A.; Lorson, C. L. *PLoS One* **2008**, *3*, e3468.

(3) Gomez-Curet, I.; Robinson, K. G.; Funanage, V. L.; Crawford, T. O.; Scavina, M.; Wang, W. *Neurogenetics* **2007**, *8*, 271.

(4) Hua, Y.; Sahashi, K.; Hung, G.; Rigo, F.; Passini, M. A.; Bennett, C. F.; Krainer, A. R. *Genes Dev.*, **2010**, *24*, 1634.

(5) Madocsai, C.; Lim, S. R.; Geib, T.; Lam, B. J.; Hertel, K. J. *Mol. Ther.* **2005**, *12*, 1013.

(6) Krainer, A. R.; Conway, G. C.; Kozak, D. *Genes Dev* **1990**, *4*, 1158.

(7) Krainer, A. R.; Conway, G. C.; Kozak, D. *Cell* **1990**, *62*, 35.

(8) Krainer, A. R.; Maniatis, T. *Cell* **1985**, *42*, 725.

(9) Krainer, A. R.; Maniatis, T.; Ruskin, B.; Green, M. R. *Cell* **1984**, *36*, 993.

(10) Ruskin, B.; Krainer, A. R.; Maniatis, T.; Green, M. R. *Cell* **1984**, *38*, 317.

(11) Mayeda, A.; Krainer, A. R. *Methods Mol. Biol.* **1999**, *118*, 315.

(12) Cartegni, L.; Hastings, M. L.; Calarco, J. A.; de Stanchina, E.; Krainer, A. R. *Am. J. Hum. Genet.* **2006**, *78*, 63.

(13) Cartegni, L.; Krainer, A. R. *Nat. Genet.* **2002**, *30*, 377.

(14) Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1987**, *15*, 8783.

(15) Nakayama, S.; Yan, L.; Sintim, H. O. *J. Am. Chem. Soc.* **2008**, *130*, 12560.

(16) Wilusz, J. E.; Devanney, S. C.; Caputi, M. *Nucleic Acids Res.* **2005**, *33*, 6547.

(17) Clarke, P. A. *Methods Mol. Biol.* **1999**, *118*, 1.

(18) Ellington, A.; Pollard, J. D., Jr. *Curr. Protoc. Mol. Biol.* **2001**, *Chapter 2*, Unit 2 12.

(19) Cartegni, L.; Krainer, A. R. *Nat. Struct. Biol.* **2003**, *10*, 120.

(20) Hicks, M. J.; Lam, B. J.; Hertel, K. J. *Methods* **2005**, *37*, 306.

(21) Hua, Y.; Vickers, T. A.; Baker, B. F.; Bennett, C. F.; Krainer, A. R. *PLoS Biol.* **2007**, *5*, e73.

(22) Hua, Y.; Vickers, T. A.; Okunola, H. L.; Bennett, C. F.; Krainer, A. R. *Am. J. Hum. Genet.* **2008**, *82*, 834.

(23) Yan, L.; Nakayama, S.; Yitbarek, S.; Greenfield, I.; Sintim, H. O. *Chem. Comm.* **2011**, *47*, 200.

(24) Kim, M. S.; Ha, C. M.; Lee, B. J. *Brain Res. Brain Res. Protoc.* **2001**, *7*, 277.